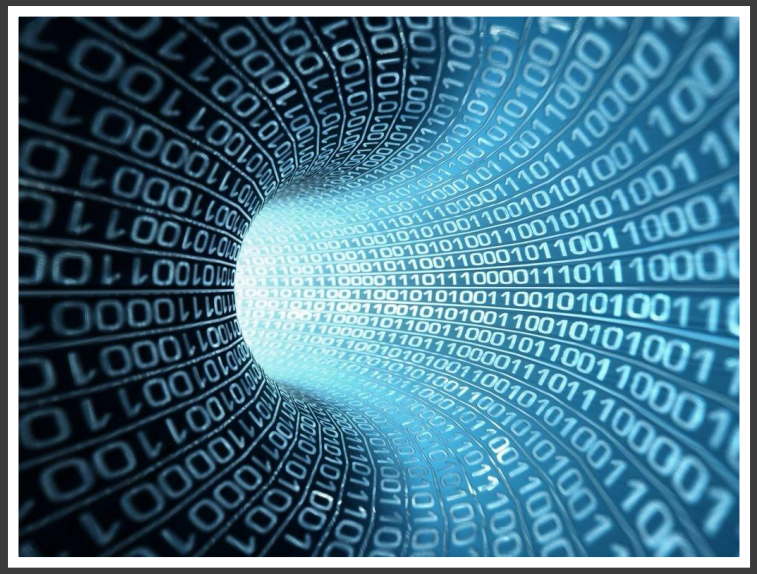


THE ROLE OF LABORATORY DATA IN 'KNOWLEDGEABLE SURVEILLANCE'

Crawford Revie (AVC / UPEI, Canada)

Fernanda Dórea (SVA, Sweden)

OUTLINE OF TALK

- **'Knowledgeable surveillance'?**
- **Recent trends in AI and machine learning**
- **Algorithms for adding value to laboratory data**
- **Integration with other data/knowledge sources**

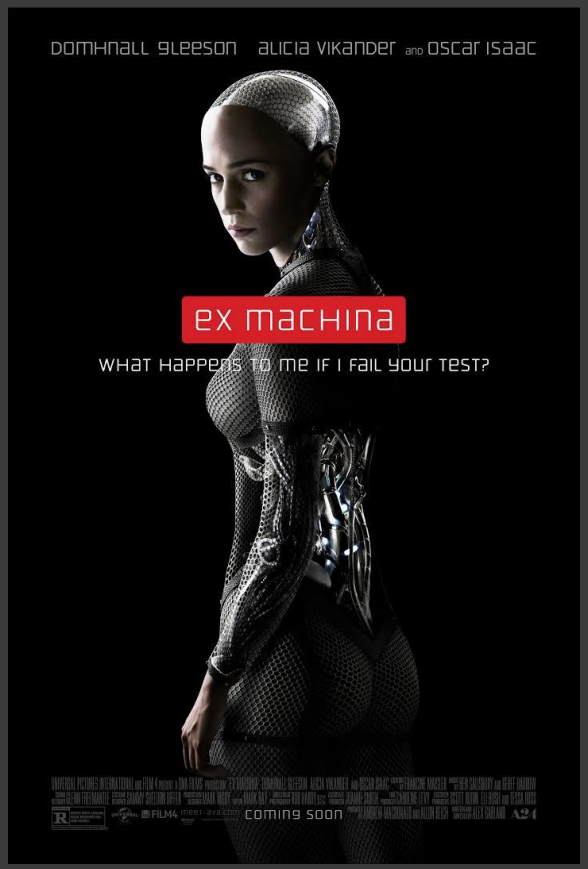
"It's what you learn after you know it all that counts."

(Harry S Truman)

knowledgeable surveillance…

epidemiological *intelligence*…

A.I.



Stephen Hawking
Stuart Russell
Max Tegmark
Frank Wilczek
(May, 2014)

Bill Gates

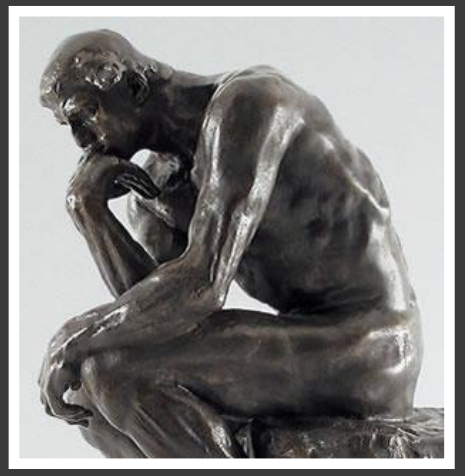
Microsoft

Elon Musk

TESLA

Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?'

WHAT HAS CHANGED?



"The fundamental tools of A.I. shifted from logic to probability in the late 1980s, and fundamental progress in the theory of uncertain reasoning underlies many of the recent practical advances."

(Peter Norvig, 2011)

IBM'S RESEARCH

Deep Blue

Watson

WATSON GETS IT WRONG

US Cities

Toronto???

gauge sentiment towards a cafe using ‘tweets‘:

- “Delicious muesli from the *@imaginarycafe* - what a <u>great</u> way to start the day!”
- “<u>Greatly</u> enjoyed the home-made lasagna at the *Imaginary Café* last night.”
- “It was a <u>great</u> disappointment to find that *Imaginary Cafe* stopped stocking BLTs.”
- “Had to wait in line for 45 minutes at the *Imaginary Cafe* today. <u>Great</u>, well there’s my lunch-break gone…”

WATSON – MORE 'SERIOUS' USES

IBM's *W*
support
worldwi
about ca

THE GLOBE AND MAIL

Search: | News & Quotes | Jobs

Enter a term, stock symbol or company name

Home | News | Opinion | Business | Investing | Sports | Life | Arts | Tec

Home » Technology

IBM's Watson to take on veterinary medicine

SHANE DINGMAN - TECHNOLOGY EDITOR
The Globe and Mail
Published Tuesday, Oct. 07, 2014 7:00PM EDT
Last updated Tuesday, Oct. 07, 2014 7:00PM EDT

1 comment | 505 | 304 | 123 | 78 | g+1 7 | Print / License | AA

International Business News

Report on

First International Business Machines Corp. took on *Jeopardy* and won with its Watson cognitive computing engine. Now, it's taking aim at animal welfare.

IBM will reveal the first 10 apps using Watson for commercial purposes on Wednesday, pushing into such areas as retail management, travel planning, medicine and computer security. Among the first apps will be Sofie, a new service made by LifeLearn Inc. of Guelph, Ont., that acts as a real-time

WHAT HAS CHANGED?

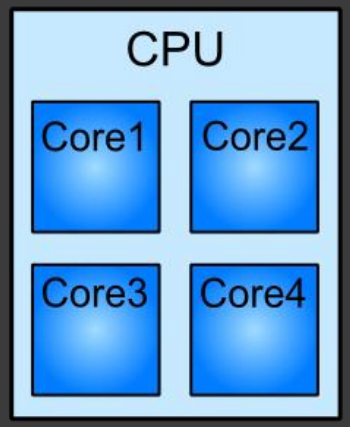
NEURAL NETWORK LAB

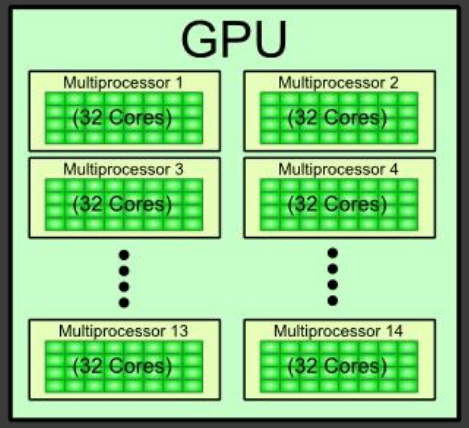
Deep Neural Networks: A Getting Started Tutorial

06/13/2014

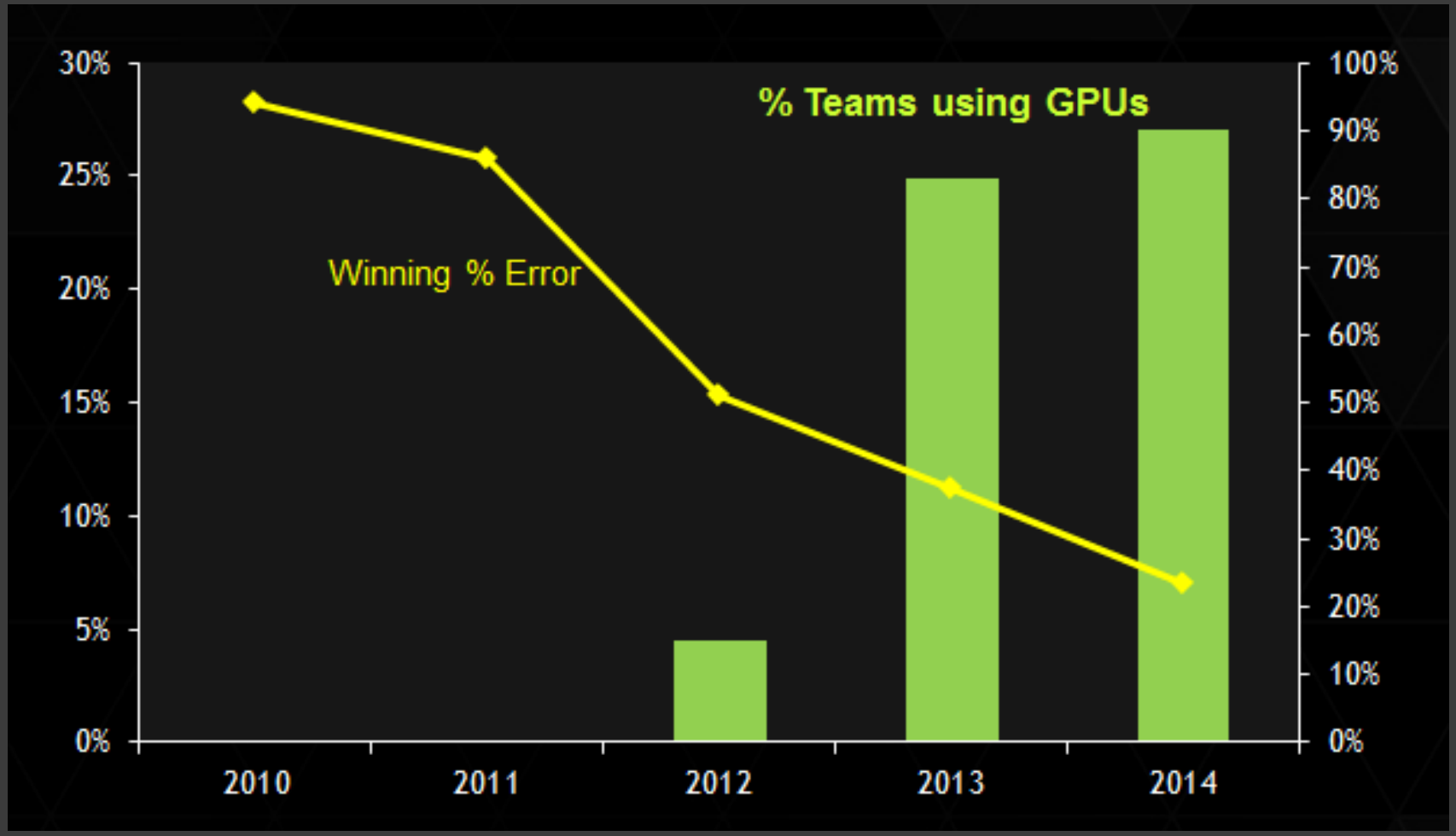
Deep Neural Networks are the more computationally powerful cousins to regular neural networks. Learn exactly what DNNs are and why they are the hottest topic in machine learning research.

Deep Learning

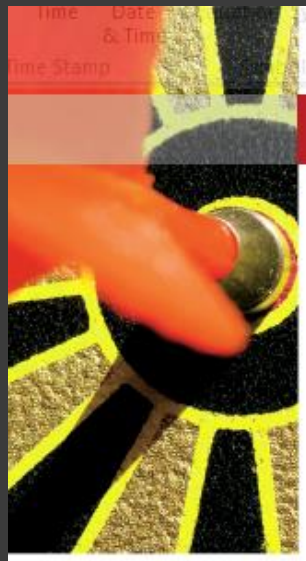




WHAT HAS CHANGED?



Results from *ILSVRC* Hinton et al (2012)
(ImageNet Large Scale Visual Recognition Challenge)



EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"[1] examines why so much of physics can be neatly explained with simple mathematical formulas such as $f = ma$ or $e = mc^2$. Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematical theories. Perhaps, unreasonable as it may seem, the right mathematical theories for human behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data.

Learning from Text at Web Scale

The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The

IEEE Intelligent Systems (2009)

The Unreasonable Effectiveness of Mathematics in the Natural Sciences. (Eugene Wigner, 1960)

MACHINE TRANSLATION

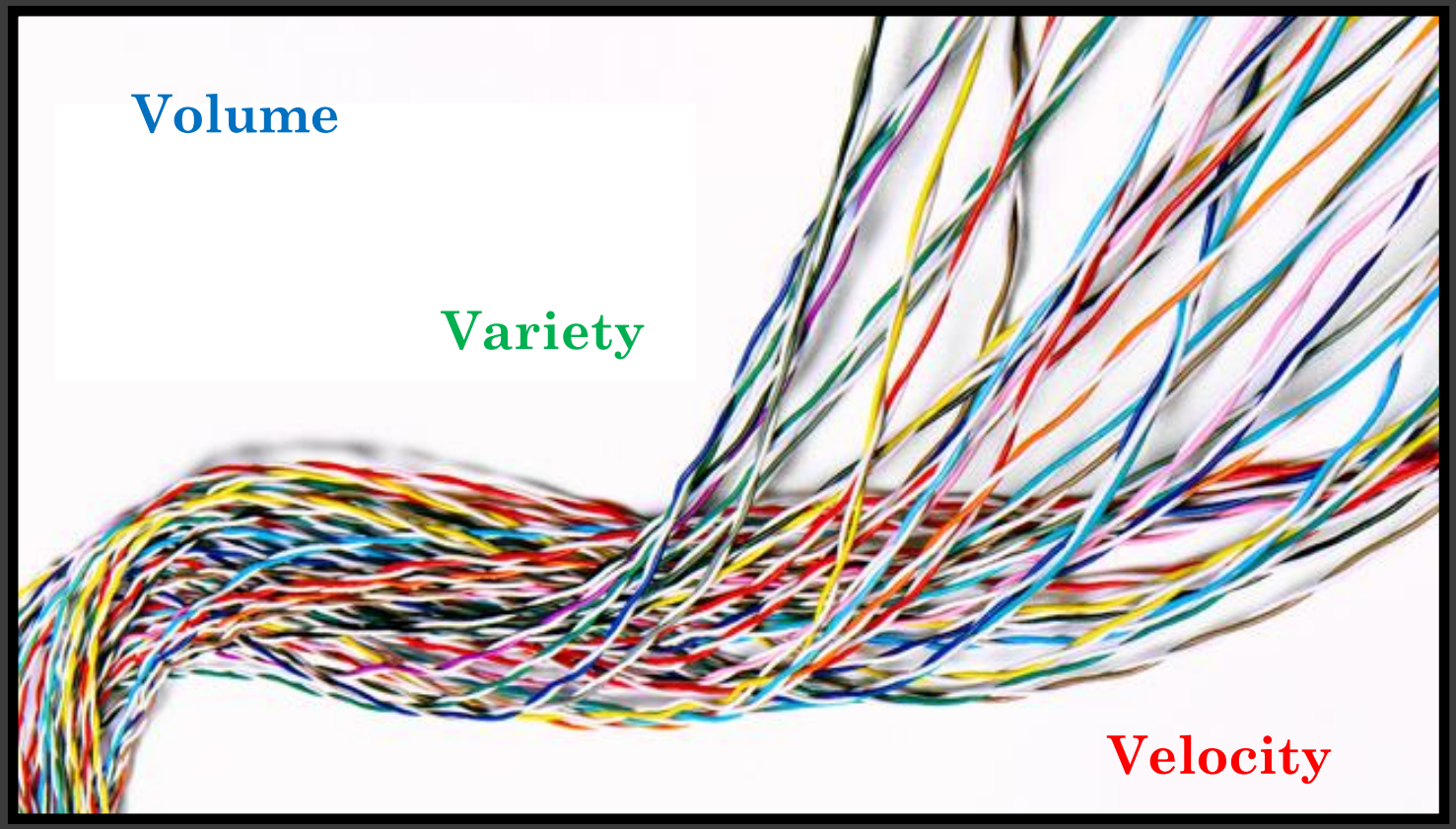
"Meanwhile, the poor Babel fish, by effectively removing all barriers to communication between different races and cultures, has caused more and bloodier wars than anything else in the history of creation."

(Hitch-Hikers Guide to the Galaxy, Douglas Adams)

DATA-DRIVEN GOVERNMENT

AND EVERYTHING ELSE…

BIG DATA



BIG DATA - VOLUME

Leslie et al (2015) The fine-scale genetic structure of the British population. *Nature*

19 March 2015

"The difference between something good and something great is attention to detail."

(Charles R Swindoll)

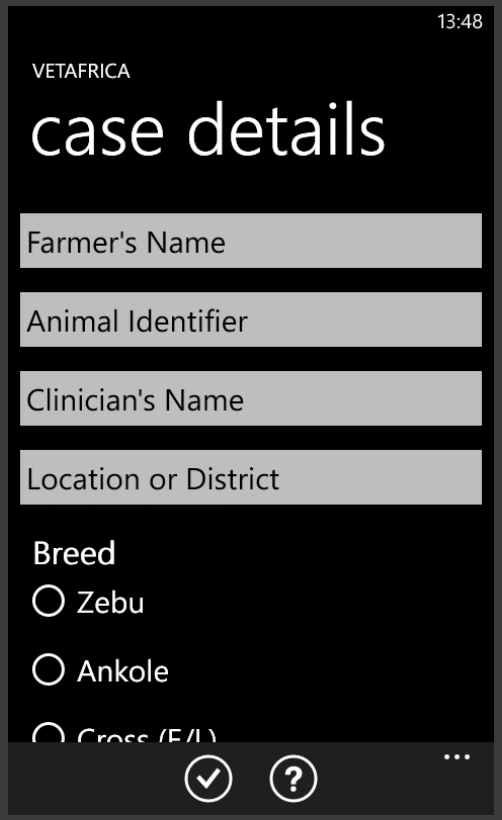
BIG DATA - VELOCITY

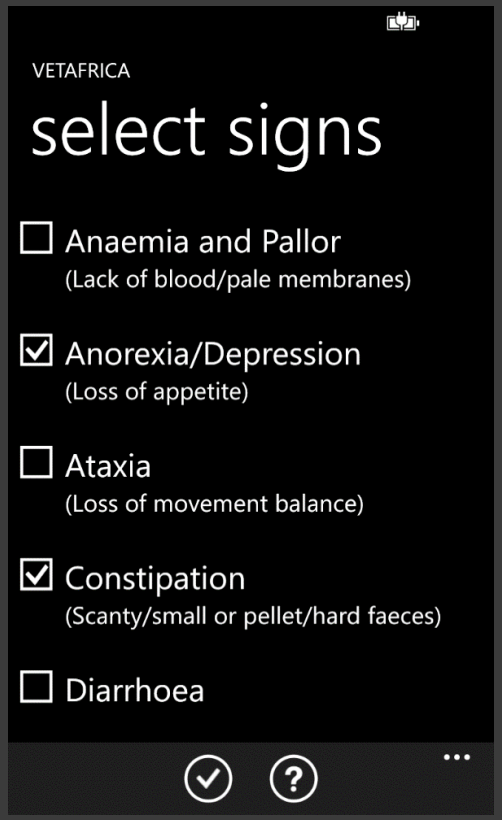
"This is the most personal device we have ever created. It's not just with you, it's on you."
(Tim Cook, Apple CEO)

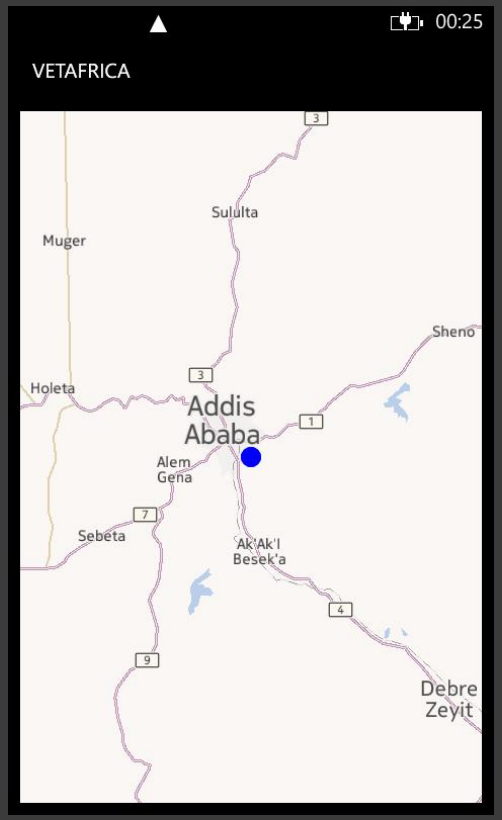
VetAfrica

BIG DATA - VARIETY

cojengo





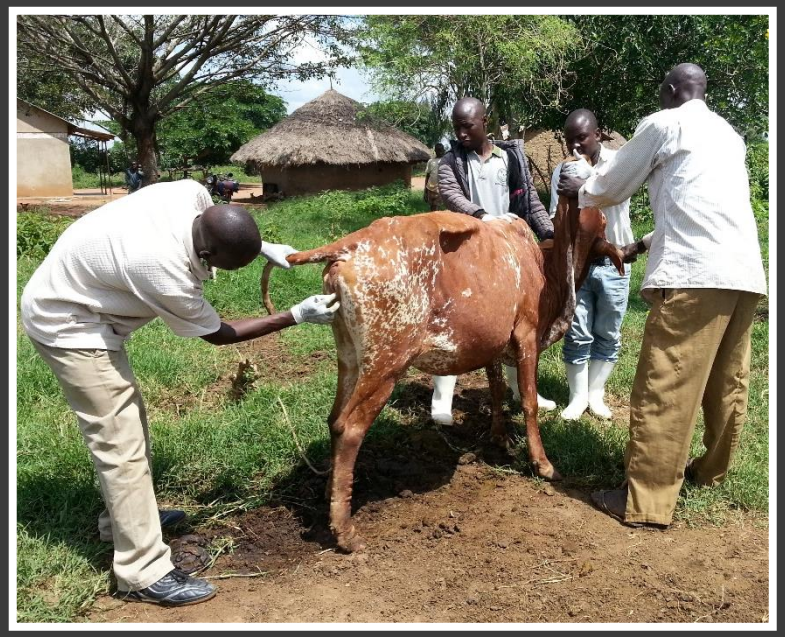


VetAfrica

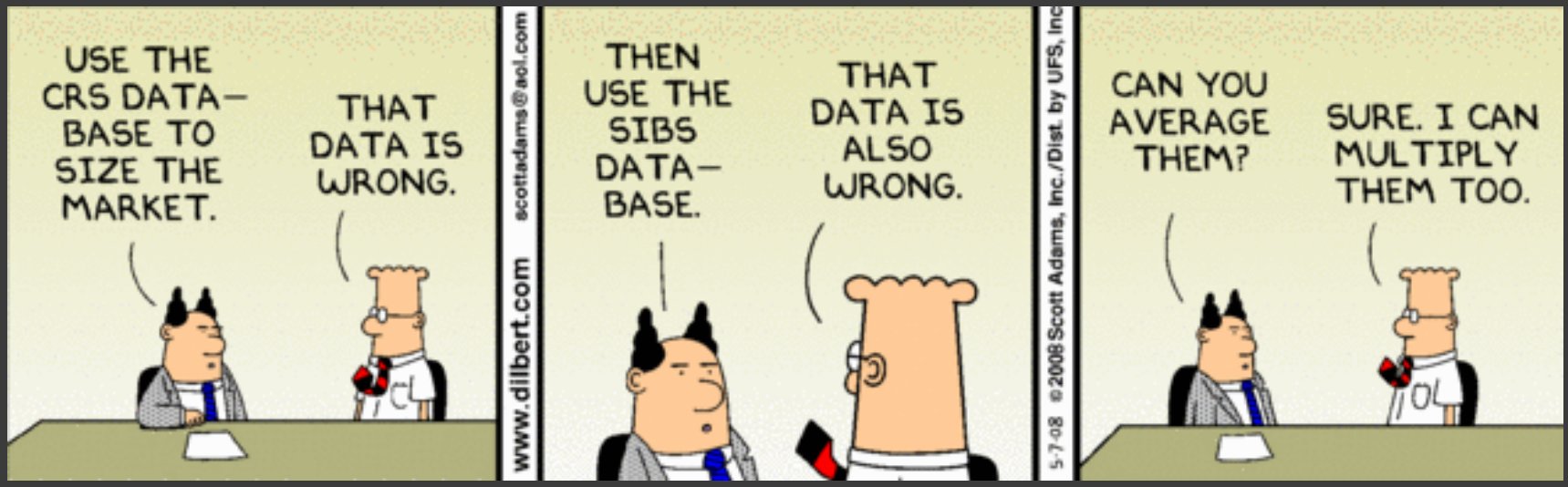
cojengo

BIG DATA - VARIETY

- **clinical signs**
- **syndromes / disease**
- **geo-spatial distribution**
- **photos of symptoms**
- **ID by photo (‘face’ recognition)**
- **body mass estimates**
- **temperature readings**
- **pen-side testing of samples**



ADDING VALUE TO LABORATORY DATA



Fancy tools/techniques are useful, but they can never compensate for poor quality data…

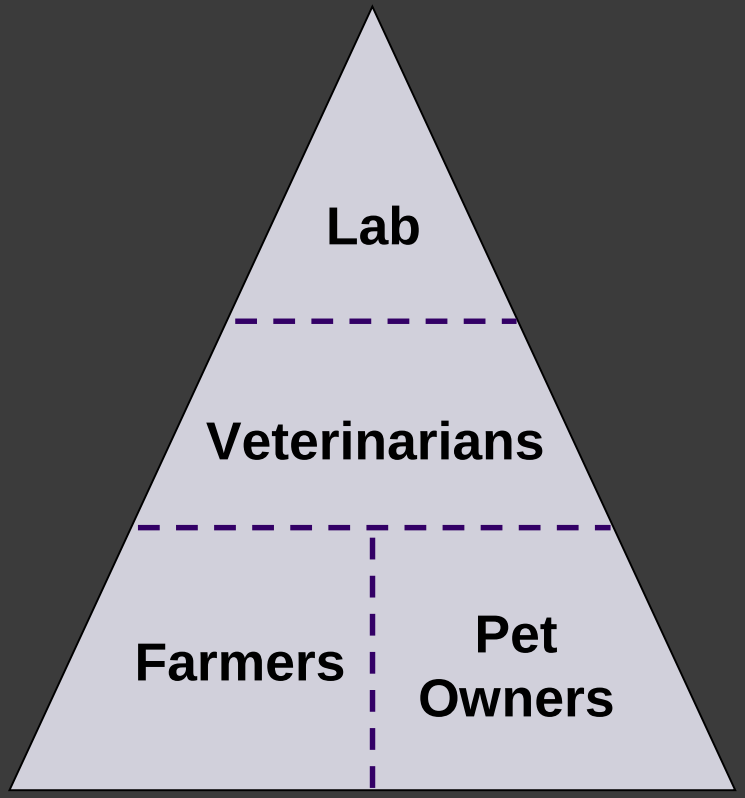
In the context of – Syndromic Surveillance

Bio-surveillance

Total Information Management

Social media / network analysis

VETERINARY SYNDROMIC SURVEILLANCE



‘Infrastructure’: pharmacies, feed companies, abattoirs, markets

Preventive Veterinary Medicine 101 (2011) 1–17

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Review

Veterinary syndromic surveillance: Current initiatives and potential for development

Fernanda C. Dórea*, Javier Sanchez, Crawford W. Revie

Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, C1A 4P3, Canada

ARTICLE INFO

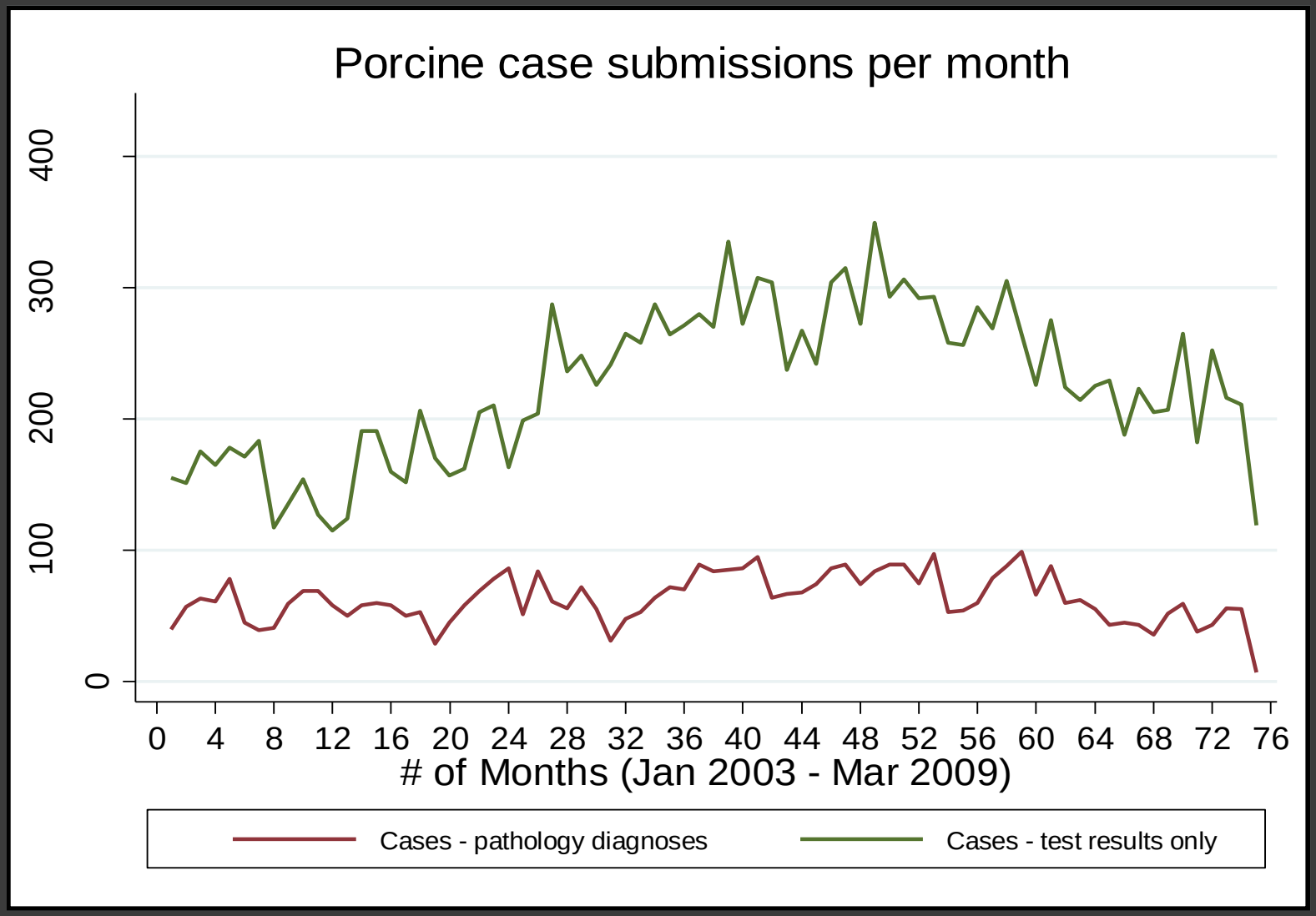
Article history:
Received 4 February 2011
Received in revised form 5 May 2011
Accepted 8 May 2011

Keywords:
Syndromic surveillance
Veterinary surveillance
Animal health surveillance
Emerging diseases
Aberration detection
Prospective monitoring

ABSTRACT

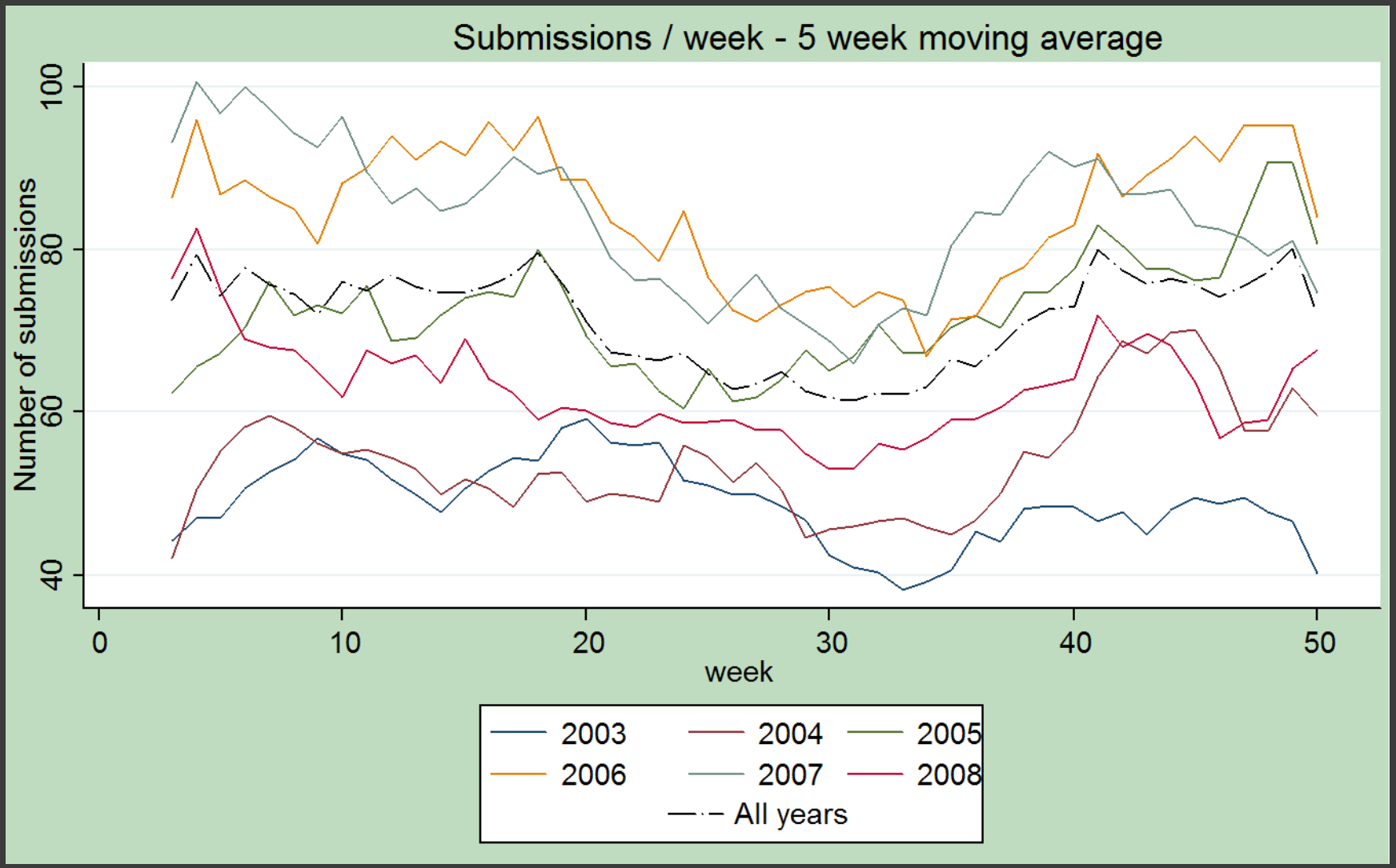
This paper reviews recent progress in the development of syndromic surveillance systems for veterinary medicine. Peer-reviewed and grey literature were searched in order to identify surveillance systems that explicitly address outbreak detection based on systematic monitoring of animal population data, in any phase of implementation. The review found that developments in veterinary syndromic surveillance are focused not only on animal health, but also on the use of animals as sentinels for public health, representing a further step towards *One Medicine*. The main sources of information are clinical data from practitioners and laboratory data, but a number of other sources are being explored. Due to limitations inherent in the way data on animal health is collected, the development of veterinary syndromic surveillance initially focused on animal health data collection strategies, analyzing historical data for their potential to support systematic monitoring, or solving problems of data classification and integration. Systems based on passive notification or data transfers are now dealing with sustainability issues. Given the ongoing barriers in availability of data, diagnostic laboratories appear to provide the most readily available data sources for syndromic surveillance in animal health. As the bottlenecks around data source availability are overcome, the next challenge is consolidating data standards for data classification, promoting the integration of different animal health surveillance systems, and also the integration to public health surveillance. Moreover, the outputs of systems for systematic monitoring of animal health data must be directly connected to real-time decision support systems which are increasingly being used for disease management and control.

WHAT DO THESE DATA LOOK LIKE?



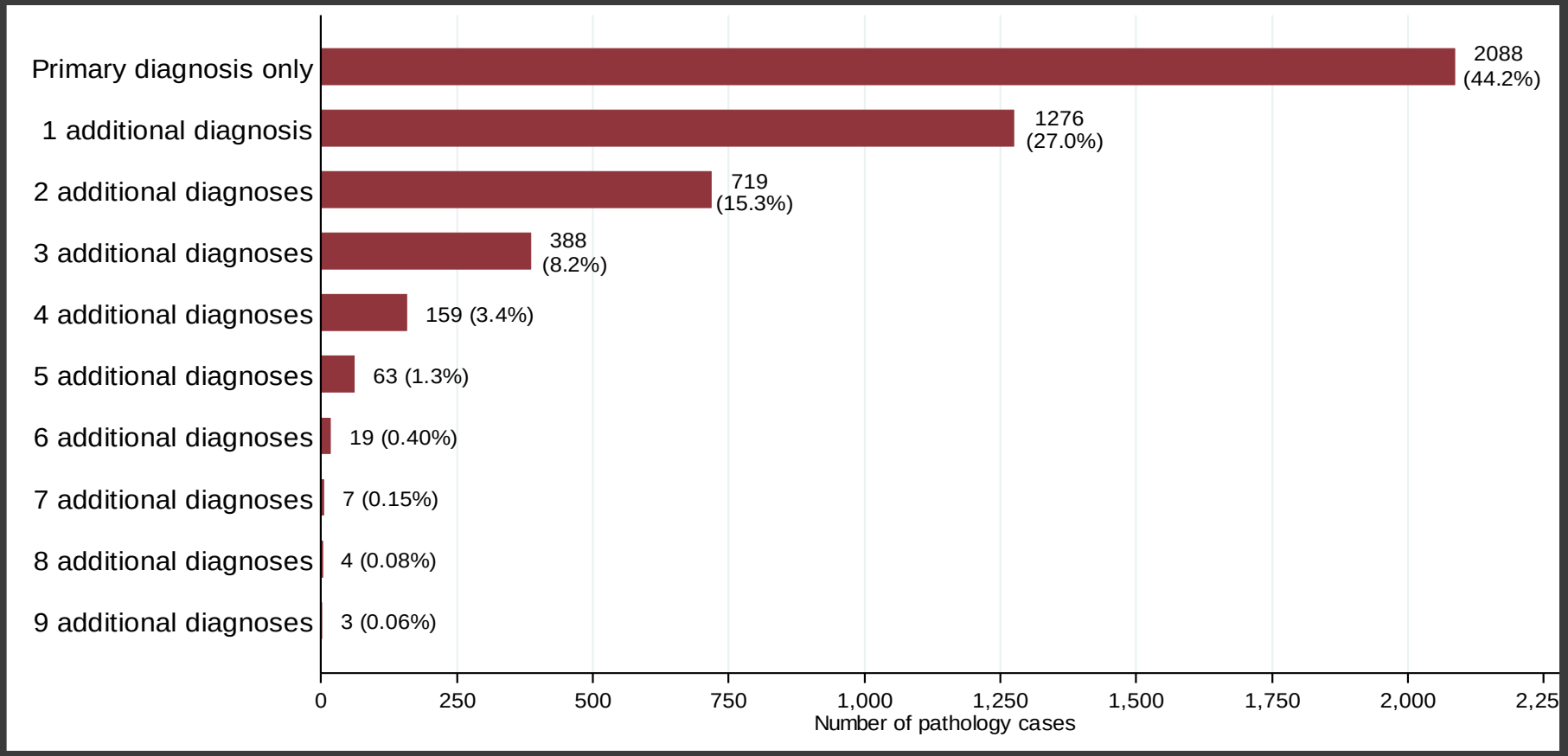
(Duizer, 2015)

WHAT DO THESE DATA LOOK LIKE?



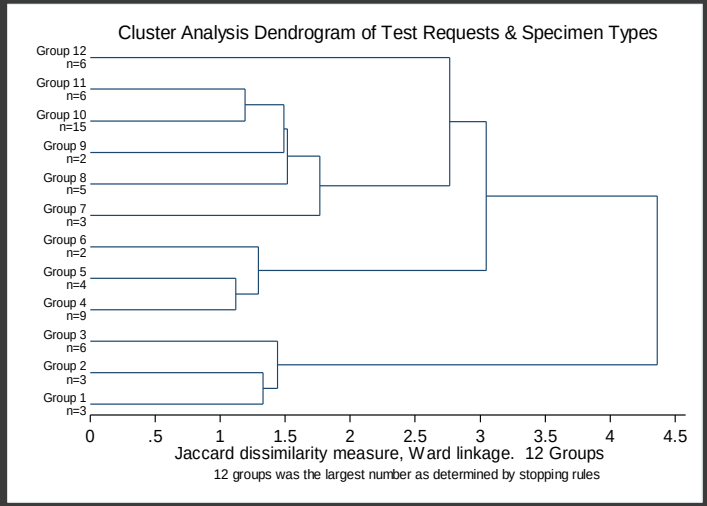
(Duizer, 2015)

DIAGNOSES PER ‘CASE’



(Duizer, 2015)

BUT WHAT IS A SYNDROME?



Group 12	FAT for Porcine parvovirus, Fetal necropsy. Fetus, fetal tissue, fetal stomach contents, female reproductive tissue
Group 11	*Streptococcus suis* typing. Neurologic tissue, brain swab, heart, spleen, kidney
Group 10	Genotyping for PCV or PRRS. Serotyping for *E coli*, PCR for Cytomegalovirus. Feces, lymph node (Not specified), pooled liver/spleen, pooled lung/spleen, tissue (Not specified), Body fluid (Not specified) Nasal swab, tonsil, serum, skin
Group 9	Micromineral analysis. Liver tissue.
Group 8	PCR for *M. hyosynoviae*. Gram stain, joint tissue, synovial membrane, joint swab
Group 7	PCR tests for *B. hyodysenteriae*, *B. pilosicoli*, *L. Intracellularis*
Group 6	PCR for *C. perfringes*. Culture tube
Group 5	FAT for Porcine corona virus (Transmissible Gastroenteritis). Fecal smear
Group 4	Electron Microscopy, anaerobic culture, ELISA for *C difficile*, *E. coli* K88 serotyping. Feces (fixed), coccidial smear. Live animal submission, all intestinal tissue
Group 3	PCR tests for swine influenza (H3N2 & H1N1), *M hyopneumoniae*, PRRS & PCV. Lung tissue
Group 2	Necropsy. Carcass, pooled lung/tonsil
Group 1	Histology, aerobic culture. Fixed tissue

(Duizer, 2015)

ALTERNATIVE VIEW OF SYNDROMES

Based on 3 years studying laboratory data sets and expert opinion

Abortion	Musculoskeletal	*Neospora caninum*	Clinical pathology
Circulatory	Nervous	Johnes disease	Toxicology
Environmental	Reproductive	BSE	MIC
Eyes and ears	Respiratory	BVD	NONSPECIFIC
GIT	Systemic		
Haematopoietic	Urinary		
Hepatic	Skin and tegument		
Mastitis			

(Dórea, 2012)

Syndrome	caseID	Source	Date	Sample	Test
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BVDV	2092582	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
Hepatic	2000032	clinical	2000-jan-02	Liver	Bacterial culture
Respiratory	2000032	clinical	2000-jan-02	Lung	Bacterial culture
Haematopoietic	2000032	clinical	2000-jan-02	Spleen	Bacterial culture
Nonspecific	2000032	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
Nonspecific	2000128	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
BVDV	2000055	clinical	2000-jan-02	Tissue_-_Fixed	BVD type 1
Nervous	2008045	clinical	2000-jan-02	Brain	Bacterial culture
Nervous	2008001	clinical	2000-jan-02	brain	Bacterial culture
Abortion	2015705	clinical	2000-jan-02	Fetal_Tissue	Bacterial Culture
Musculoeskeletal	2015789	clinical	2000-jan-02	bone	Clostridia
Musculoeskeletal	2015789	clinical	2000-jan-02	joint	Clostridia
Respiratory	2015719	clinical	2000-jan-02	Lung	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
Respiratory	2015705	clinical	2000-jan-02	Bronchoalveolar_Lavage	Bacterial culture
GIT	2024811	clinical	2000-jan-02	large Int	Clostridium perfringens
Respiratory	2024699	clinical	2000-jan-02	Lung	Virus isolation

CAN WE AUTOMATICALLY CLASSIFY?

Direct mapping:

- Based on keywords – 25% of all records

Text mining algorithms:

- Dictionary of relevant words
- Direct relationships
- Co-occurrences
- Precedence

CLIENTSAMPLEID	SAMPLETYPEDESC	PARAMLISTDESC	PARAMLISTID
90	Blood - Serum	Bovine leukemia virus - ELISA	BLV ELISA 2 step
90	Blood - Serum	Bovine viral diarrhea virus, type 1 Singer - VN	BVD 1 Singer VN
90	Blood - Serum	Bovine herpesvirus 1 (Infectious bovine rhinotracl	IBR VN
90	Blood - Serum	Sample split	Serum Split
300	Blood - Serum	Bovine leukemia virus - ELISA	BLV ELISA 2 step
300	Blood - Serum	Mycobacterium paratuberculosis, bovine - ELISA	M paratb ELISA
300	Blood - Serum	Bovine herpesvirus 1 (Infectious bovine rhinotracl	IBR VN
300	Blood - Serum	Bovine viral diarrhea virus, type 1 Singer - VN	BVD 1 Singer VN
300	Blood - Serum	Sample split	Serum Split
436	Blood - Serum	Mycobacterium paratuberculosis, bovine - ELISA	M paratb ELISA
20	Serum	Biochemistry profile, bovine	Bovine Profile
20	Blood - EDTA	Complete blood count, comprehensive	CBC
Olive	Serum	Bovine leukemia virus - ELISA	BLV ELISA 2 step
85	Blood - Serum	Bovine leukemia virus - ELISA	BLV ELISA 2 step
54	Serum	Mineral panel, trace elements, serum	M panel, serum
78	Serum	Mineral panel, trace elements, serum	M panel, serum
95	Serum	Mineral panel, trace elements, serum	M panel, serum
162	Serum	Mineral panel, trace elements, serum	M panel, serum

CAN WE AUTOMATICALLY CLASSIFY?

	Class average (Macro)		
Algorithm	**Recall**	**Precision**	**F-score**
Manually modified rules	0.994	1.000	0.997
Naïve Bayes	0.983	0.939	0.955
Rule Induction	0.626	0.793	0.677
Decision Trees	0.290	0.416	0.311

(Dórea et al, 2012)

Records classification

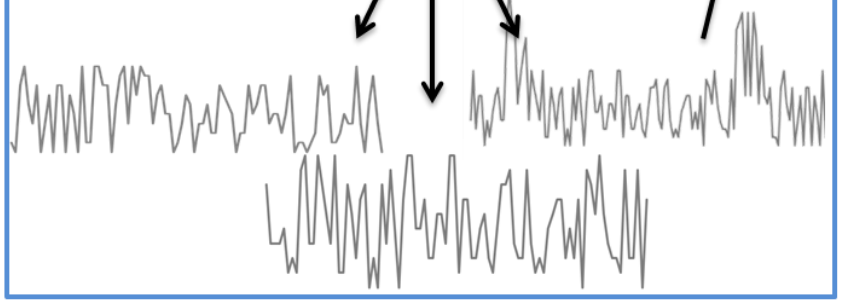
Data extraction

caseID	Source	Date	Sample	Test
2092814	clinical	2000-jan-01	Serum	BLV
2092814	clinical	2000-jan-01	Serum	BLV
2092582	clinical	2000-jan-01	Serum	BVD type 2
2092607	clinical	2000-jan-01	Serum	BVD type 2
2092607	clinical	2000-jan-01	Serum	BVD type 2
2000032	clinical	2000-jan-02	Liver	Bacterial culture
2000032	clinical	2000-jan-02	Lung	Bacterial culture
2000032	clinical	2000-jan-02	Spleen	Bacterial culture
2000032	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
2000128	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
2000055	clinical	2000-jan-02	Tissue_-_Fixed	BVD type 1
2008045	clinical	2000-jan-02	Brain	Bacterial culture
2008001	clinical	2000-jan-02	brain	Bacterial culture
2015705	clinical	2000-jan-02	Fetal_Tissue	Bacterial Culture
2015789	clinical	2000-jan-02	bone	Clostridia
2015789	clinical	2000-jan-02	joint	Clostridia
2015719	clinical	2000-jan-02	Lung	Bacterial culture
2015705	clinical	2000-jan-02	rumen	Bacterial culture
2015705	clinical	2000-jan-02	rumen	Bacterial culture
2015705	clinical	2000-jan-02	Bronchoalveolar_Lavage	Bacterial culture
2024811	clinical	2000-jan-02	large Int	Clostridium perfringens
2024699	clinical	2000-jan-02	Lung	Virus isolation

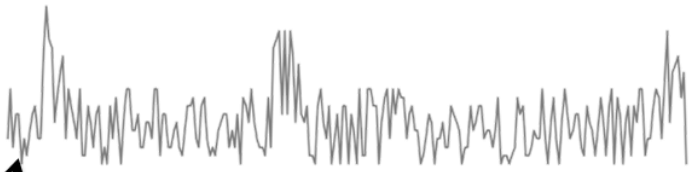
Classification of records into syndromes

Syndrome	caseID	Source	Date	Sample	Test
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BVDV	2092582	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
Hepatic	2000032	clinical	2000-jan-02	Liver	Bacterial culture
Respiratory	2000032	clinical	2000-jan-02	Lung	Bacterial culture
Haematopoietic	2000032	clinical	2000-jan-02	Spleen	Bacterial culture
Nonspecific	2000032	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
Nonspecific	2000128	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
BVDV	2000055	clinical	2000-jan-02	Tissue_-_Fixed	BVD type 1
Nervous	2008045	clinical	2000-jan-02	Brain	Bacterial culture
Nervous	2008001	clinical	2000-jan-02	brain	Bacterial culture
Abortion	2015705	clinical	2000-jan-02	Fetal_Tissue	Bacterial Culture
Musculoeskeletal	2015789	clinical	2000-jan-02	bone	Clostridia
Musculoeskeletal	2015789	clinical	2000-jan-02	joint	Clostridia
Respiratory	2015719	clinical	2000-jan-02	Lung	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
Respiratory	2015705	clinical	2000-jan-02	Bronchoalveolar_Lavage	Bacterial culture
GIT	2024811	clinical	2000-jan-02	large Int	Clostridium perfringens
Respiratory	2024699	clinical	2000-jan-02	Lung	Virus isolation

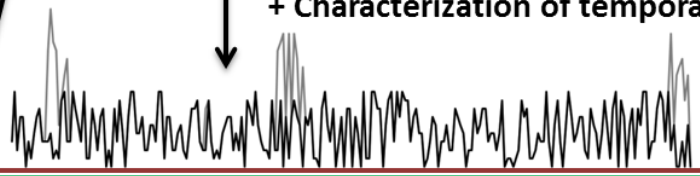
Case counts for each syndrome



Retrospective analysis

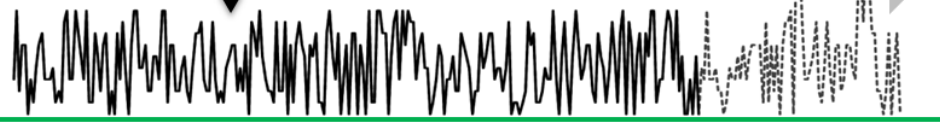


Removal of aberrations and excessive noise
+ Characterization of temporal effects



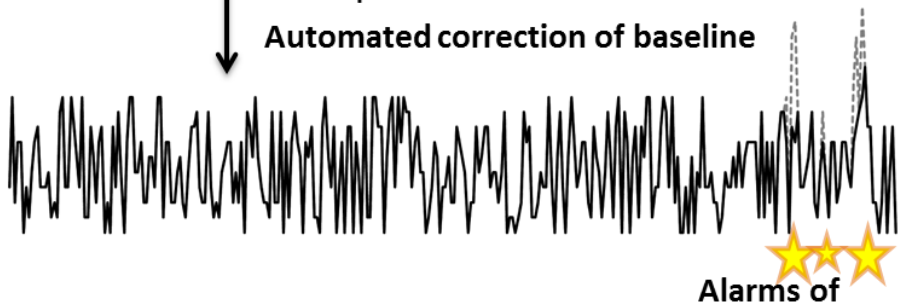
Prospective analysis

Use historical data as a baseline to train algorithms for prospective detection of aberrations



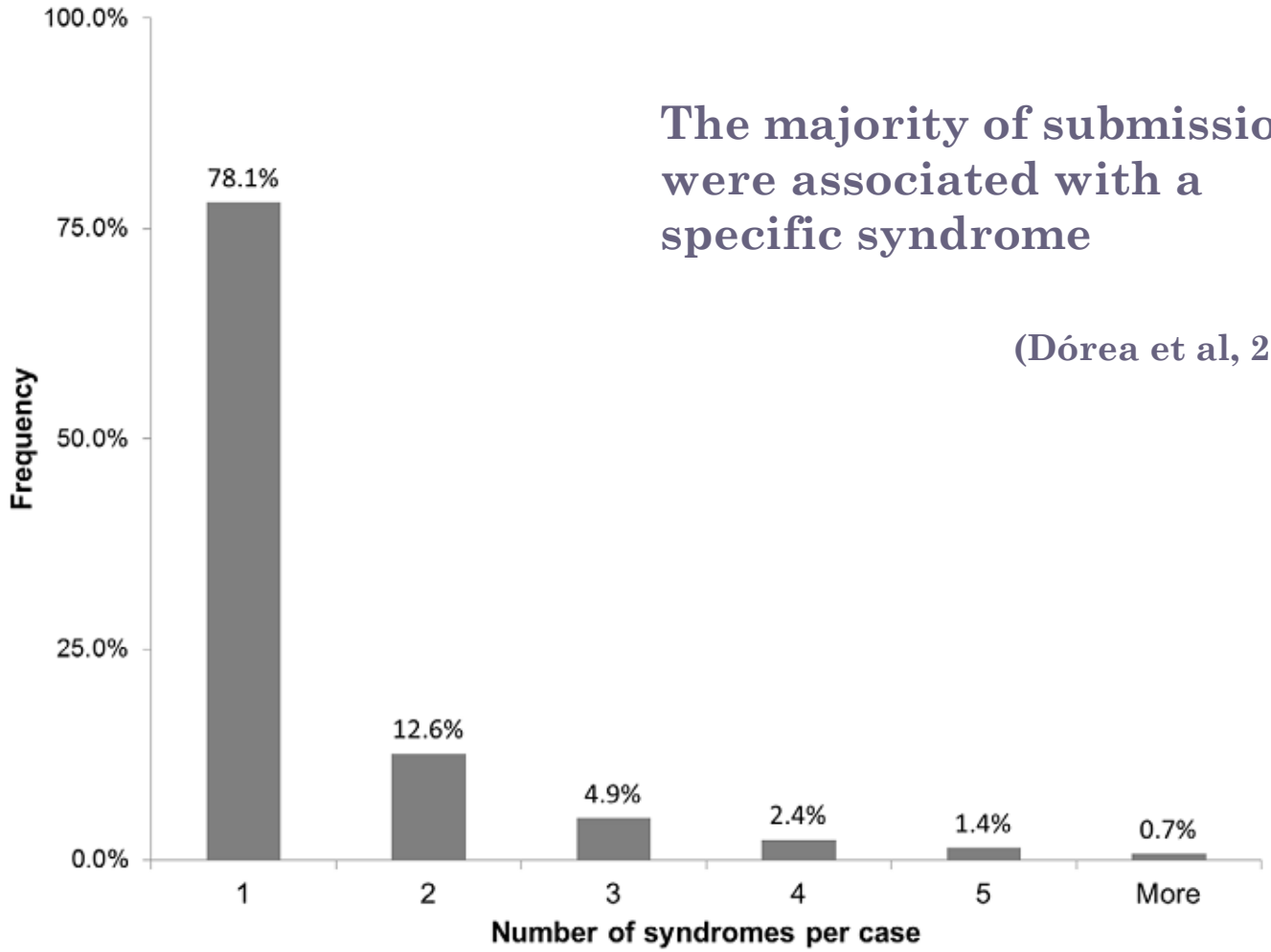
Real-time monitoring

Alarm detection
+
Automated correction of baseline



Alarms of various magnitudes

CASE DEFINITION

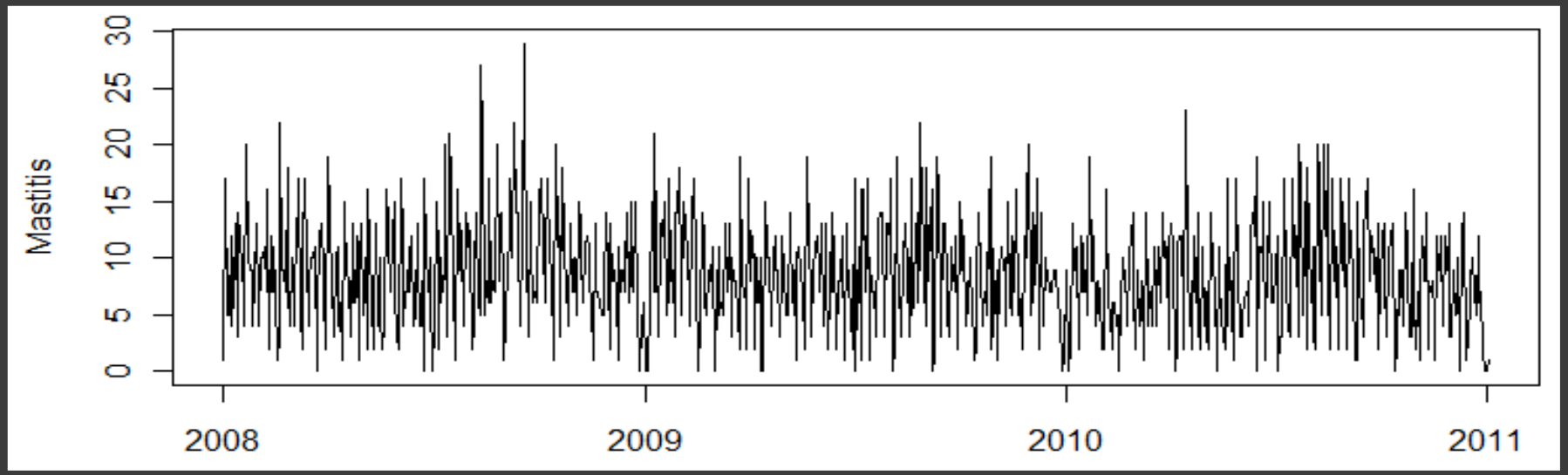


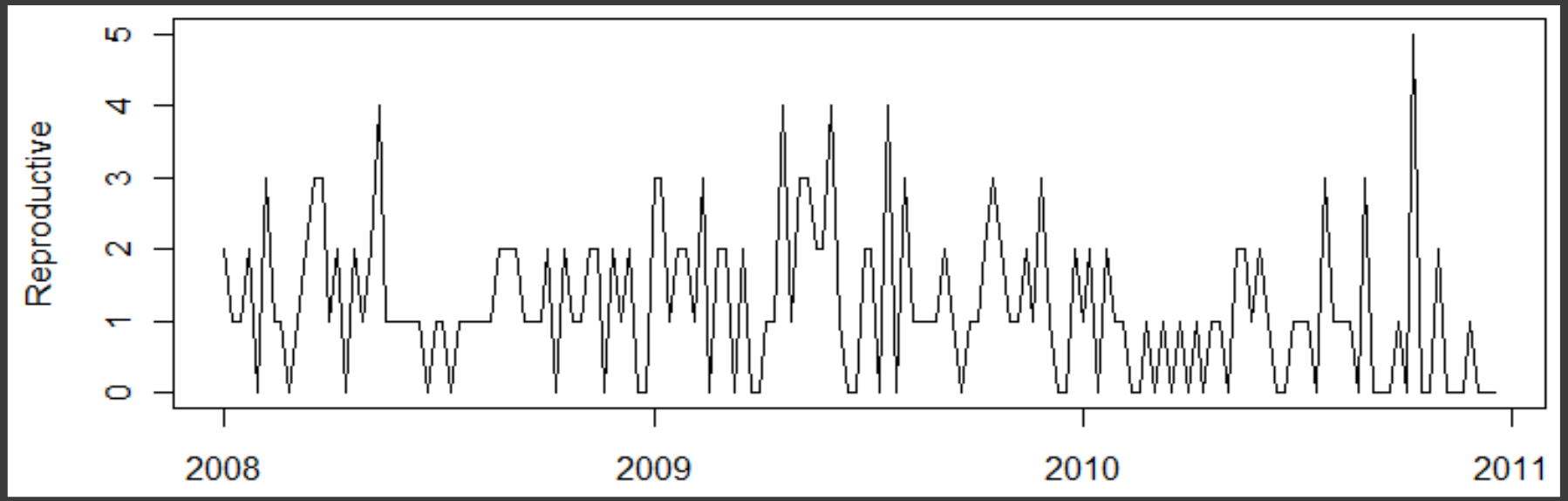
The majority of submissions were associated with a specific syndrome

(Dórea et al, 2013)

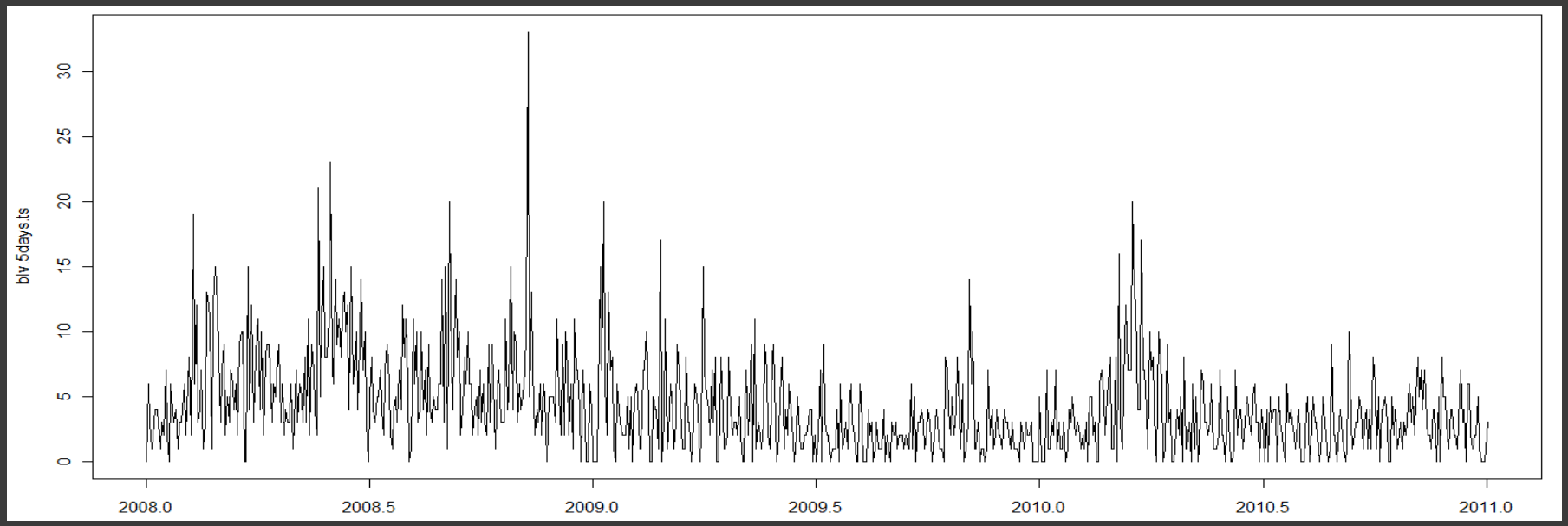
Figure 1. Number of syndromes identified in each case using information from individual test requests.
doi:10.1371/journal.pone.0057334.g001

VARIATION IN SYNDROME FREQUENCY



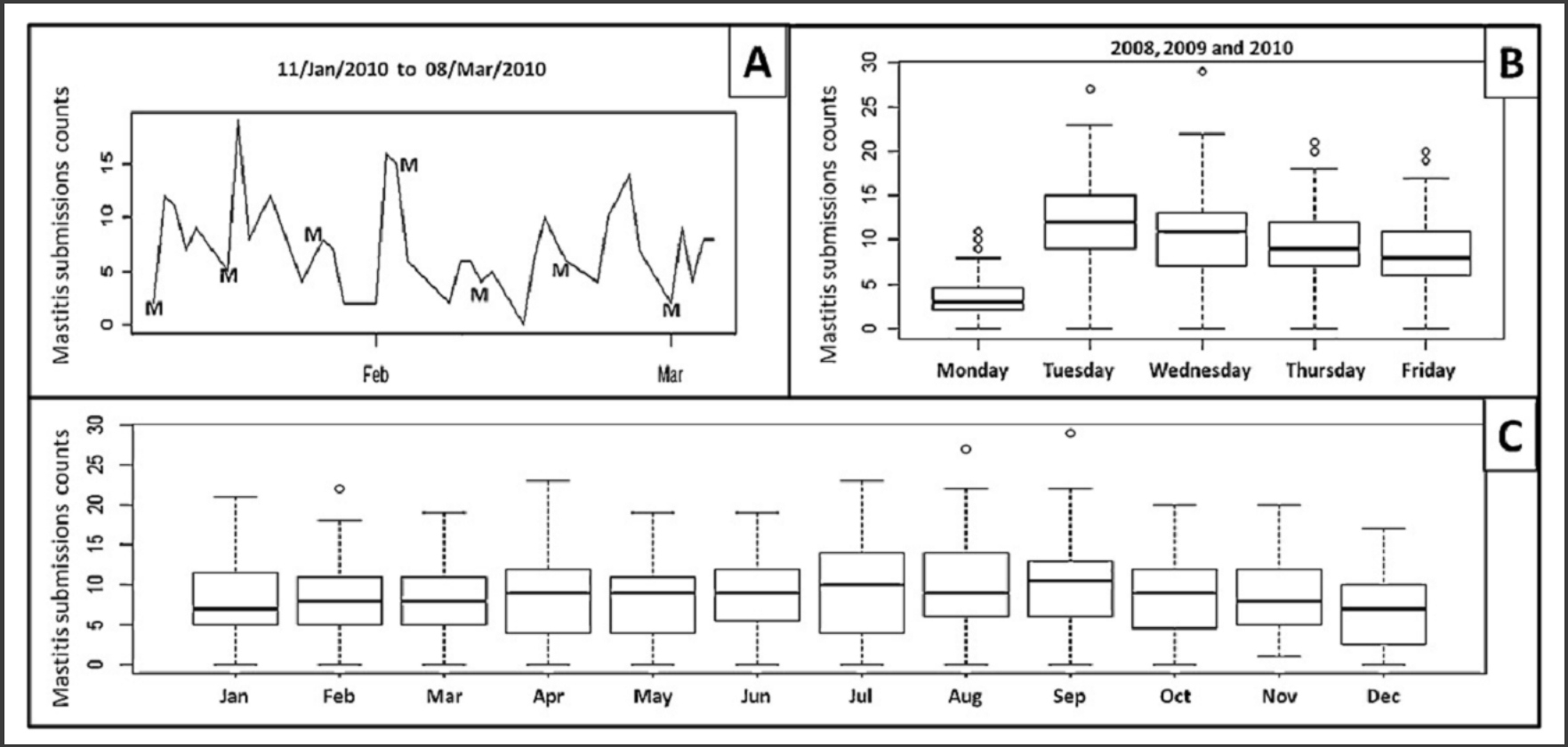


BASELINE - RETROSPECTIVE ANALYSIS



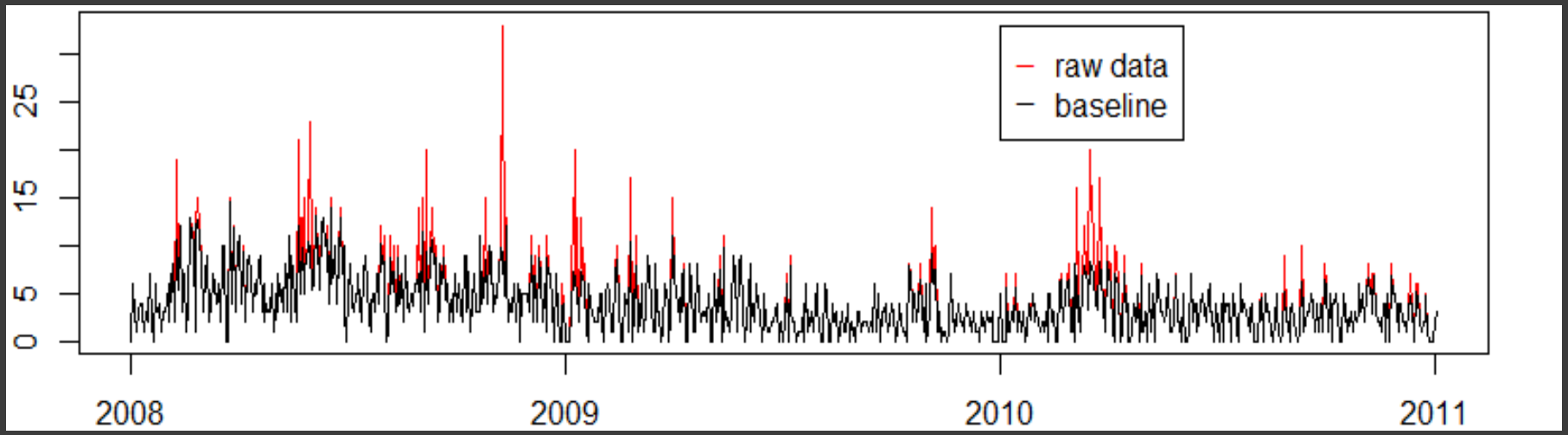
Noise versus past outbreaks
Small counts per day
Temporal effects

BASELINE - RETROSPECTIVE ANALYSIS



Day of the week (DOW)
Monthly variation (explicit vs Sine fct.?)

BASELINE - RETROSPECTIVE ANALYSIS



Poisson regression model with DOW and monthly effects for BLV at daily submission.

(Dórea et al, 2012)

PROSPECTIVE ANALYSIS AS A MEANS TO ABERRATION DETECTION

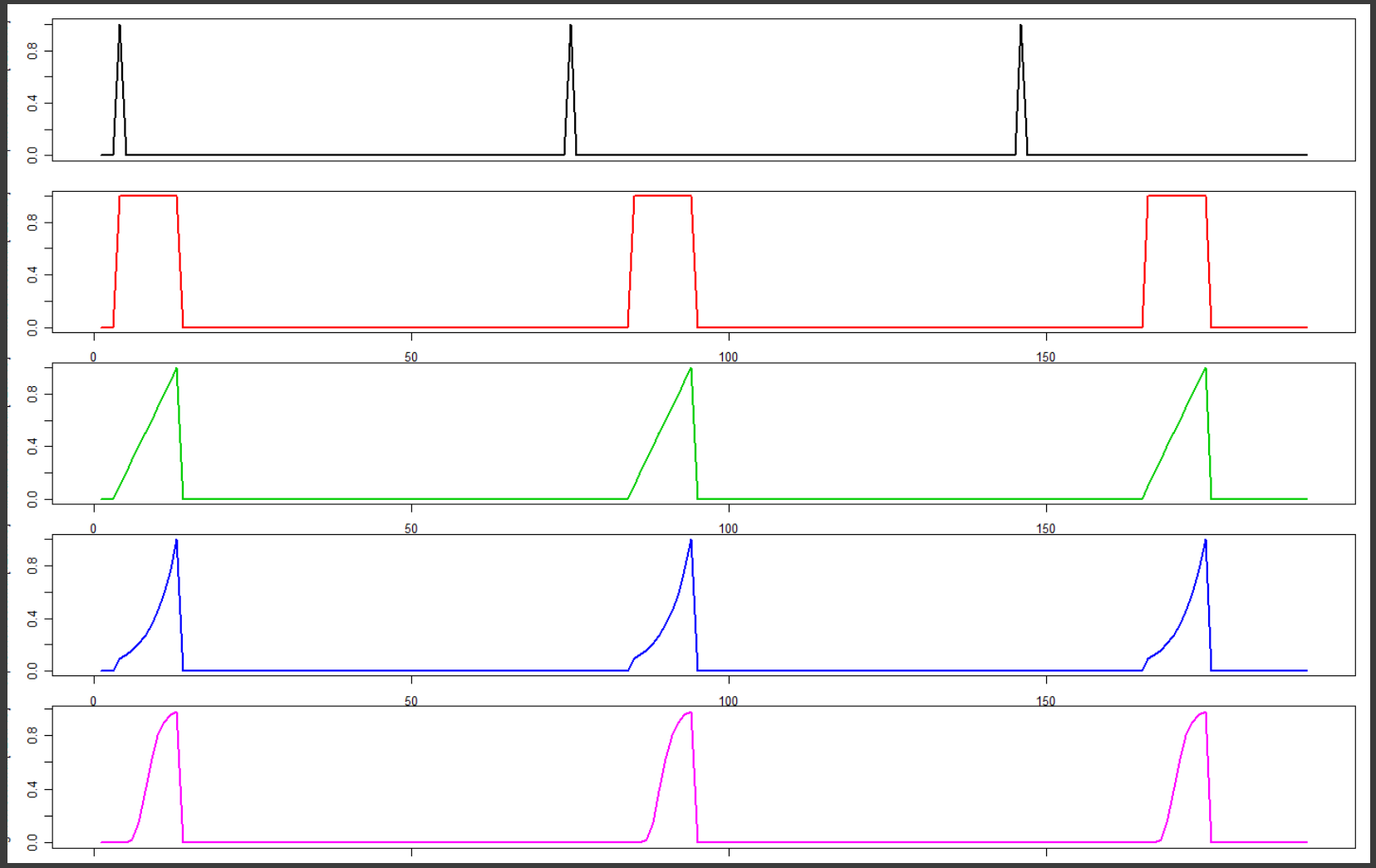
Which algorithms?

- Control charts
  - Shewhart
  - Cumulative sums (CUSUM)
  - Exponentially Weighted Moving Averages (EWMA)
- Regression models + Control chart
  - Poisson, Negative Binomial
- Differencing + Control charts
- Holt-Winters exponential smoothing

Characteristic of the outbreak

Characteristics of the data stream

BUT WHAT MIGHT AN OUTBREAK LOOK LIKE?



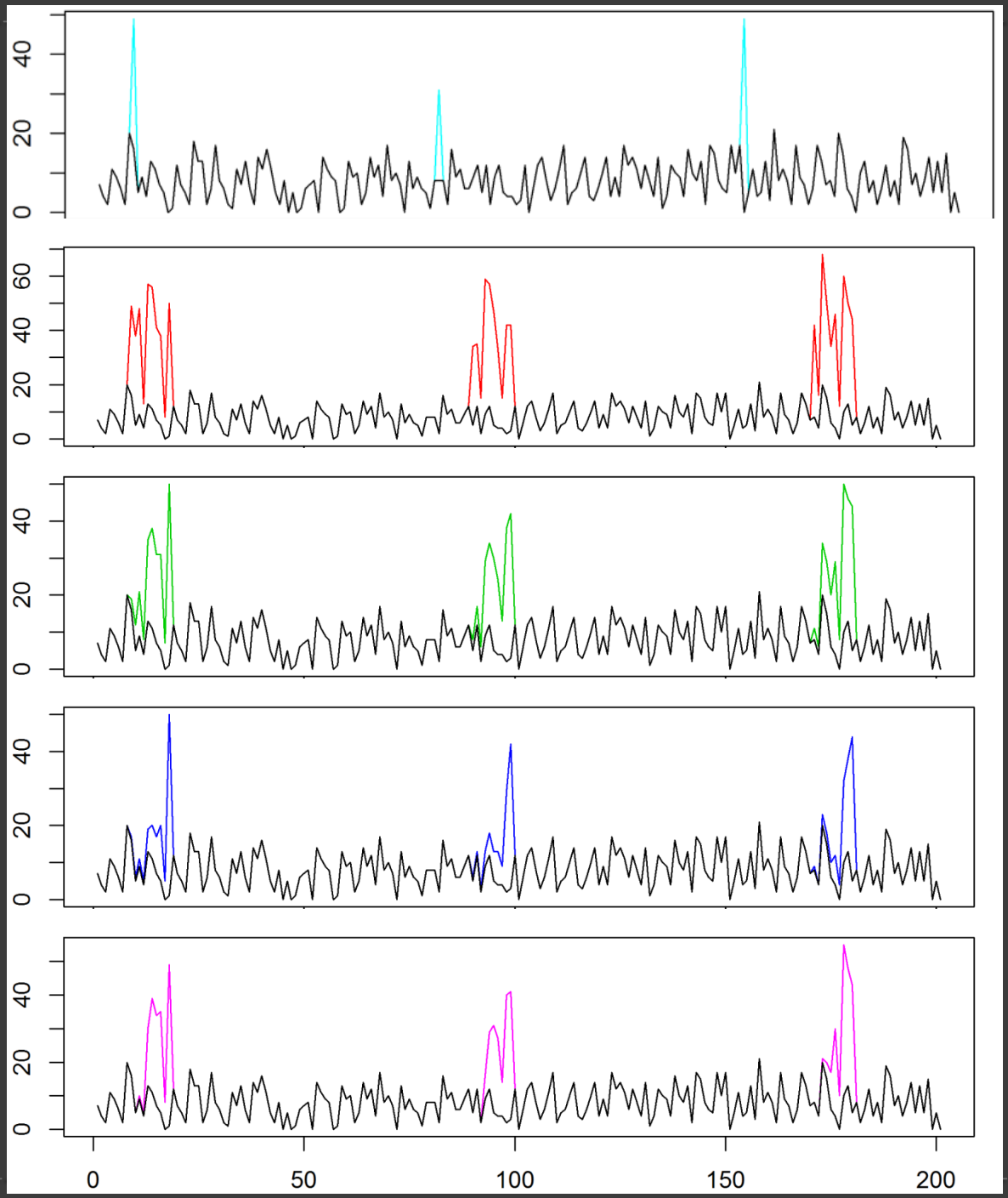
Spike

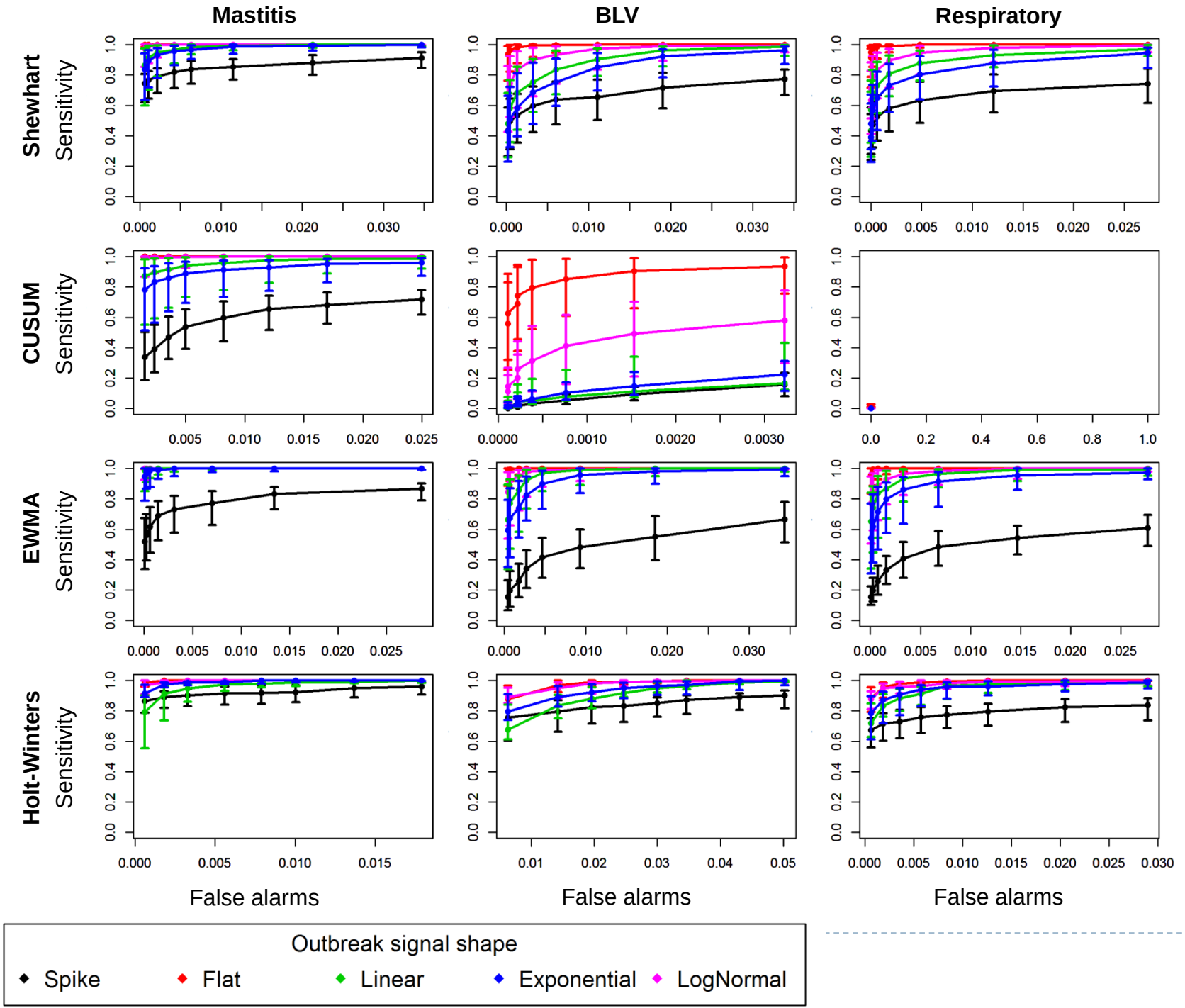
Flat

Linear

Exponential

Log-Normal




Mastitis
BLV
Respiratory
Shewhart
CUSUM
EWMA
Holt-Winters
Sensitivity
False alarms
Outbreak signal shape
Spike
Flat
Linear
Exponential
LogNormal

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

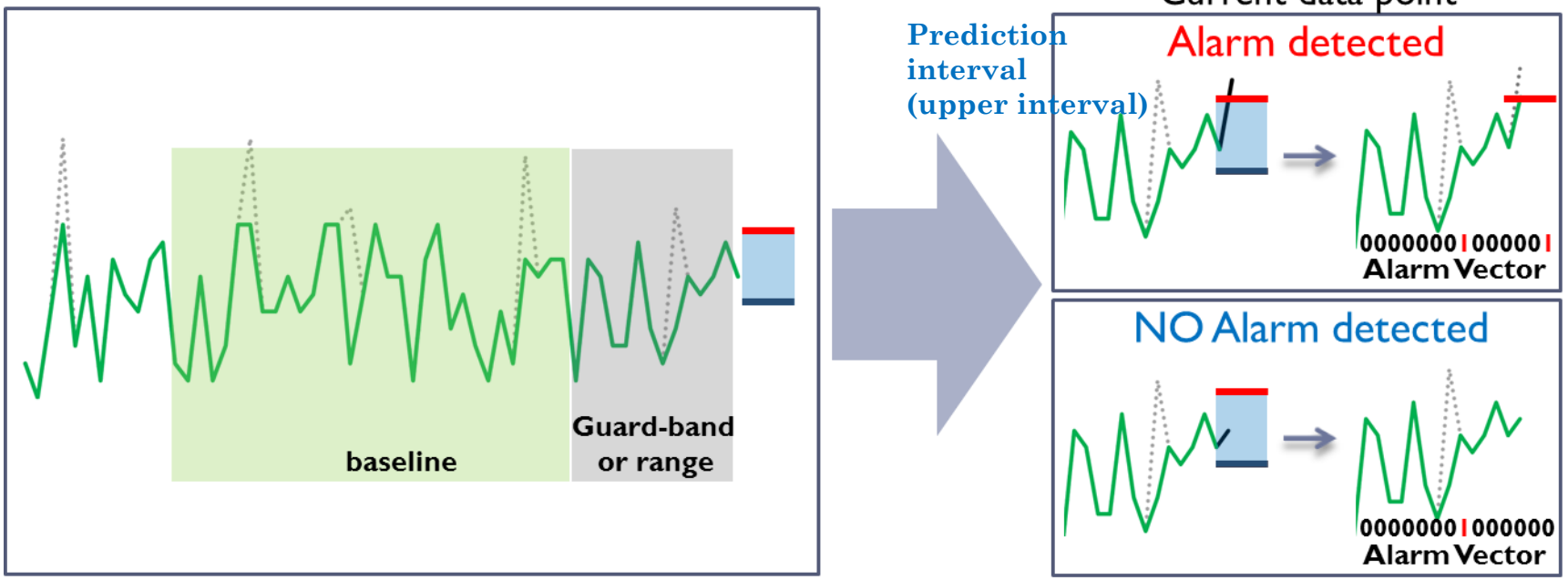
Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation

Fernanda C. Dórea[1], Beverly J. McEwen[2], W. Bruce McNab[3], Crawford W. Revie[1] and Javier Sanchez[1]

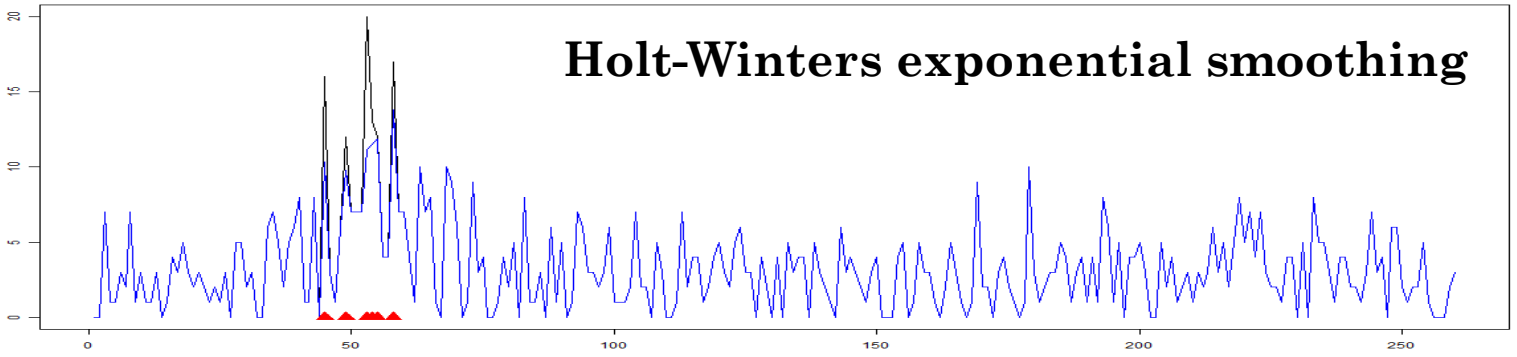
[1]Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada C1A 4P3
[2]Animal Health Laboratory, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1H 6R8
[3]Ontario Ministry of Agriculture Food and Rural Affairs, Guelph, Ontario, Canada N1G 4Y2

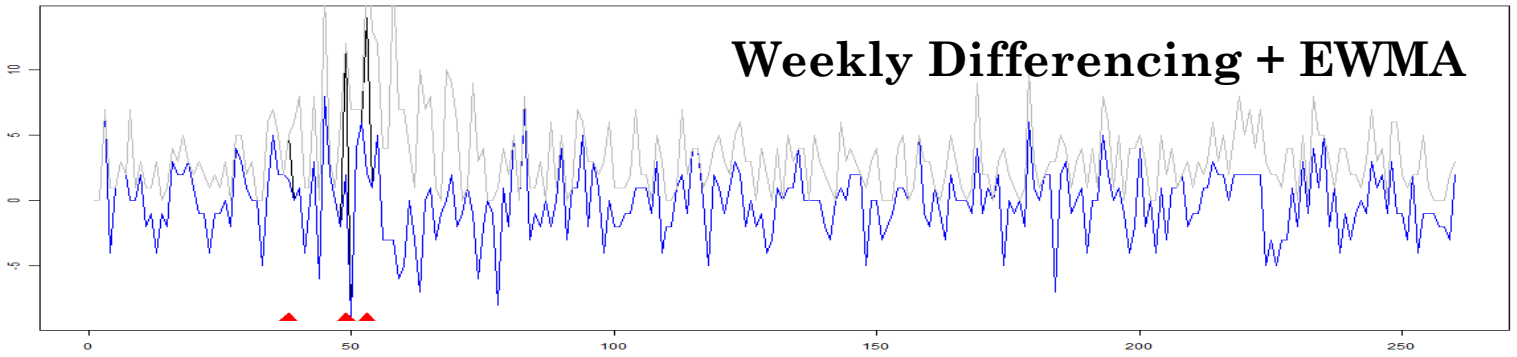
(Dórea et al, 2013)

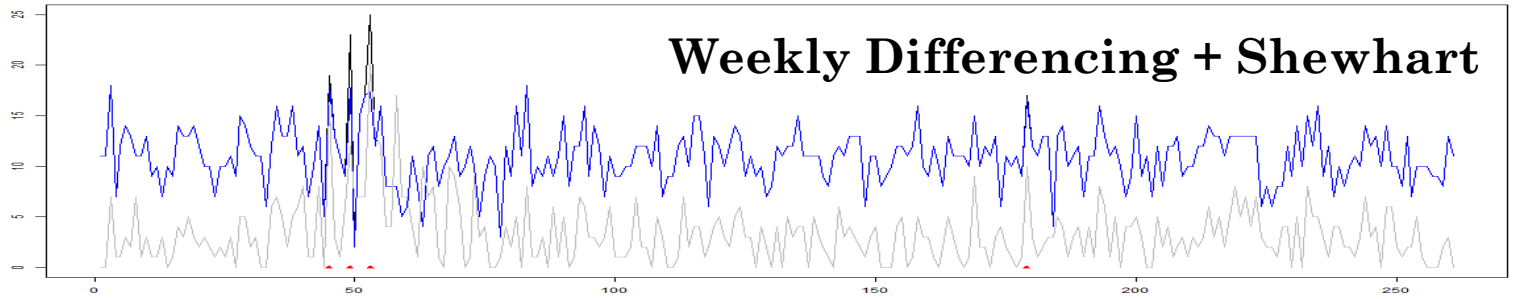
IMPLEMENTATION: ABERRATION CORRECTION AND INTERPRETATION



IMPLEMENTATION: ABERRATION CORRECTION AND INTERPRETATION

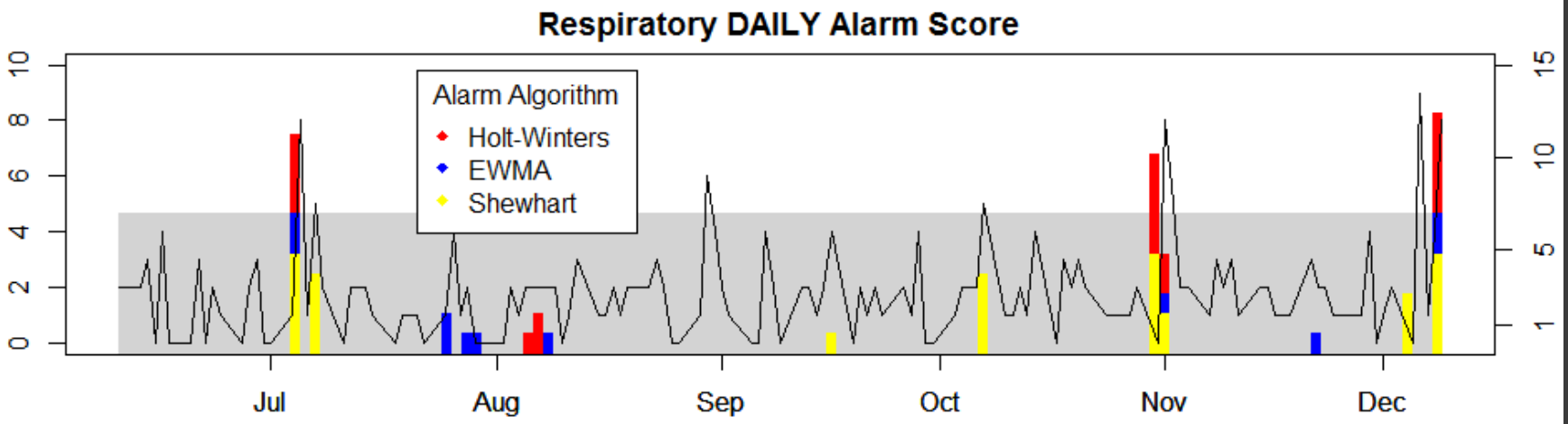


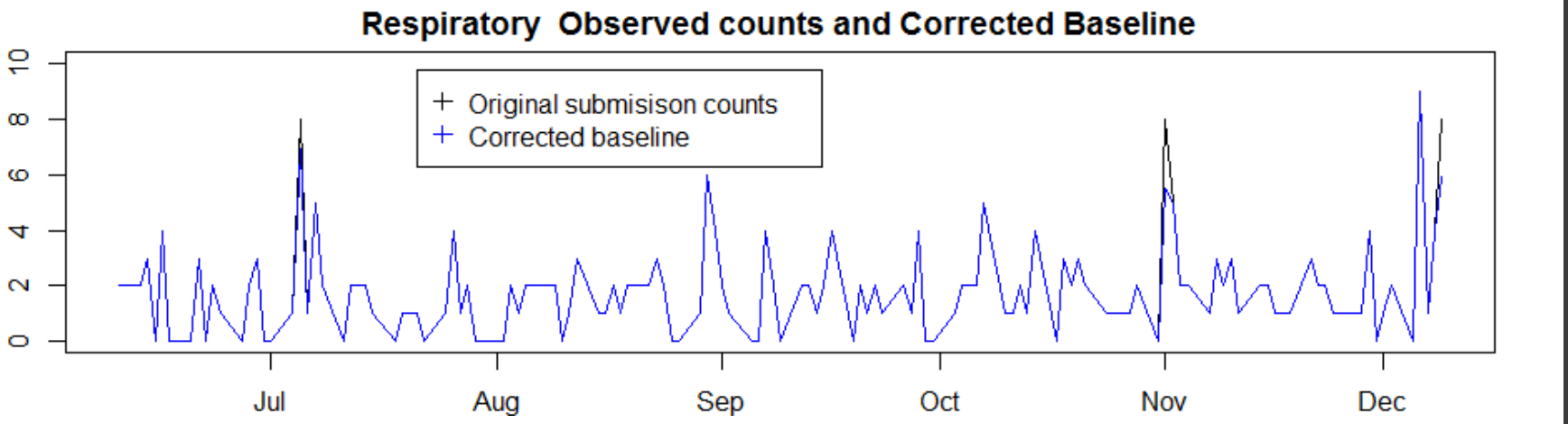




Respiratory - 2011-12-09

Algorithm	D-4	D-3	D-2	D-1	TODAY
HW	0	0	0	0	5
EWMA	0	0	0	0	2
Shewhart	0	3	0	0	5
SUM	0	3	0	0	12





Records classification

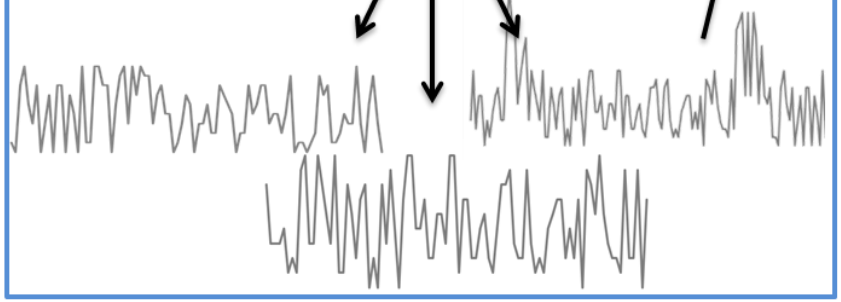
Data extraction

caseID	Source	Date	Sample	Test
2092814	clinical	2000-jan-01	Serum	BLV
2092814	clinical	2000-jan-01	Serum	BLV
2092582	clinical	2000-jan-01	Serum	BVD type 2
2092607	clinical	2000-jan-01	Serum	BVD type 2
2092607	clinical	2000-jan-01	Serum	BVD type 2
2000032	clinical	2000-jan-02	Liver	Bacterial culture
2000032	clinical	2000-jan-02	Lung	Bacterial culture
2000032	clinical	2000-jan-02	Spleen	Bacterial culture
2000032	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
2000128	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
2000055	clinical	2000-jan-02	Tissue_-_Fixed	BVD type 1
2008045	clinical	2000-jan-02	Brain	Bacterial culture
2008001	clinical	2000-jan-02	brain	Bacterial culture
2015705	clinical	2000-jan-02	Fetal_Tissue	Bacterial Culture
2015789	clinical	2000-jan-02	bone	Clostridia
2015789	clinical	2000-jan-02	joint	Clostridia
2015719	clinical	2000-jan-02	Lung	Bacterial culture
2015705	clinical	2000-jan-02	rumen	Bacterial culture
2015705	clinical	2000-jan-02	rumen	Bacterial culture
2015705	clinical	2000-jan-02	Bronchoalveolar_Lavage	Bacterial culture
2024811	clinical	2000-jan-02	large Int	Clostridium perfringens
2024699	clinical	2000-jan-02	Lung	Virus isolation

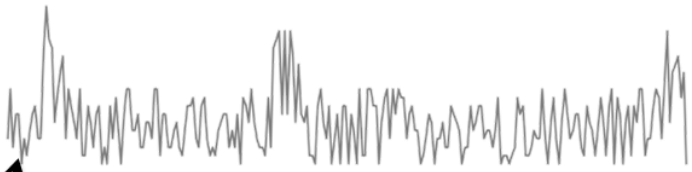
Classification of records into syndromes

Syndrome	caseID	Source	Date	Sample	Test
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BLV	2092814	clinical	2000-jan-01	Serum	BLV
BVDV	2092582	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
BVDV	2092607	clinical	2000-jan-01	Serum	BVD type 2
Hepatic	2000032	clinical	2000-jan-02	Liver	Bacterial culture
Respiratory	2000032	clinical	2000-jan-02	Lung	Bacterial culture
Haematopoietic	2000032	clinical	2000-jan-02	Spleen	Bacterial culture
Nonspecific	2000032	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
Nonspecific	2000128	clinical	2000-jan-02	Tissue_-_Fixed	Histopathology
BVDV	2000055	clinical	2000-jan-02	Tissue_-_Fixed	BVD type 1
Nervous	2008045	clinical	2000-jan-02	Brain	Bacterial culture
Nervous	2008001	clinical	2000-jan-02	brain	Bacterial culture
Abortion	2015705	clinical	2000-jan-02	Fetal_Tissue	Bacterial Culture
Musculoeskeletal	2015789	clinical	2000-jan-02	bone	Clostridia
Musculoeskeletal	2015789	clinical	2000-jan-02	joint	Clostridia
Respiratory	2015719	clinical	2000-jan-02	Lung	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
GIT	2015705	clinical	2000-jan-02	rumen	Bacterial culture
Respiratory	2015705	clinical	2000-jan-02	Bronchoalveolar_Lavage	Bacterial culture
GIT	2024811	clinical	2000-jan-02	large Int	Clostridium perfringens
Respiratory	2024699	clinical	2000-jan-02	Lung	Virus isolation

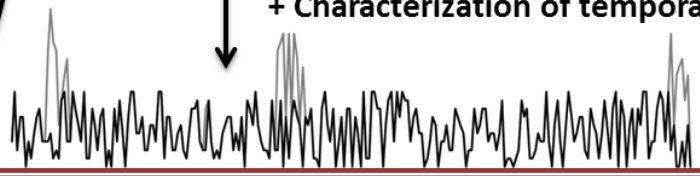
Case counts for each syndrome



Retrospective analysis

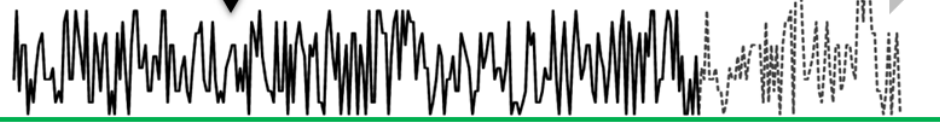


Removal of aberrations and excessive noise
+ Characterization of temporal effects



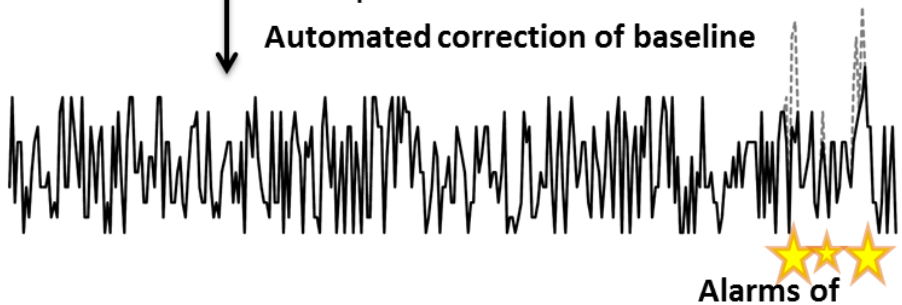
Prospective analysis

Use historical data as a baseline to train algorithms for prospective detection of aberrations



Real-time monitoring

Alarm detection
+
Automated correction of baseline



Alarms of various magnitudes

Development of an *R package for syndromic surveillance*: translating research into tools for the veterinary epidemiologist

Fernanda C. Dórea[1], Stefan Widgrén[1], Crawford W. Revie[2], Ann Lindberg[1]

1- Swedish Zoonosis Centre. Department of Disease Control and Epidemiology. National Veterinary Institute (SVA), Sweden.

2- Department of Health Management. Atlantic Veterinary College. University of Prince Edward Island, Canada.

Sweden

Canada

Centre for Veterinary Epidemiological Research

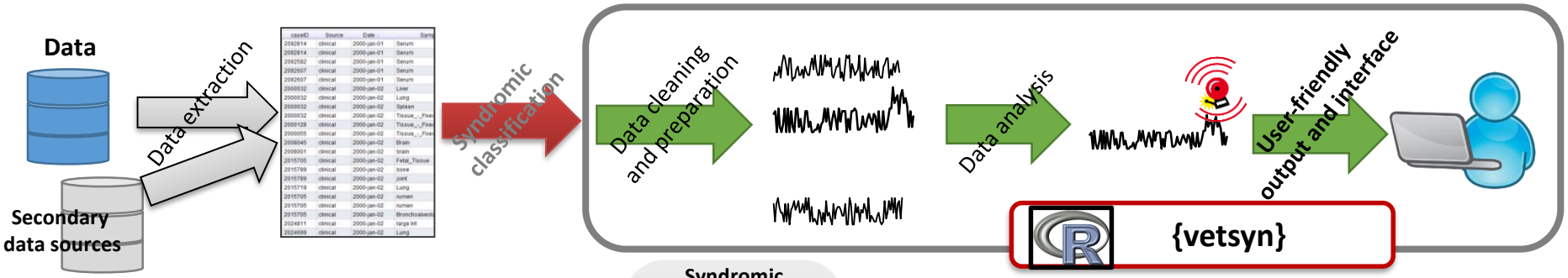
Abstract

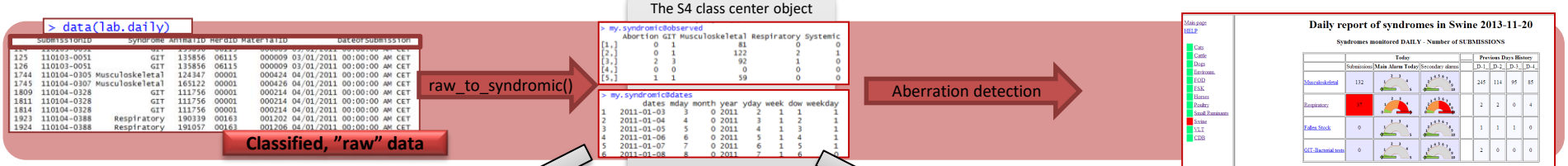
We describe the development of an R package for veterinary syndromic surveillance. Initial development has focused on the use of health data already classified into syndromic groups. Algorithms are available to conduct retrospective analysis of data and create outbreak-free baselines that can serve as training sets. In addition, a range of aberration detection algorithms have been proven to efficiently detect outbreak signals in animal health data streams of various magnitudes (median number of observations per day from one to hundreds) and containing various temporal effects. Functions are available to easily develop an html interface for such systems and set up automated emails in case of alarms

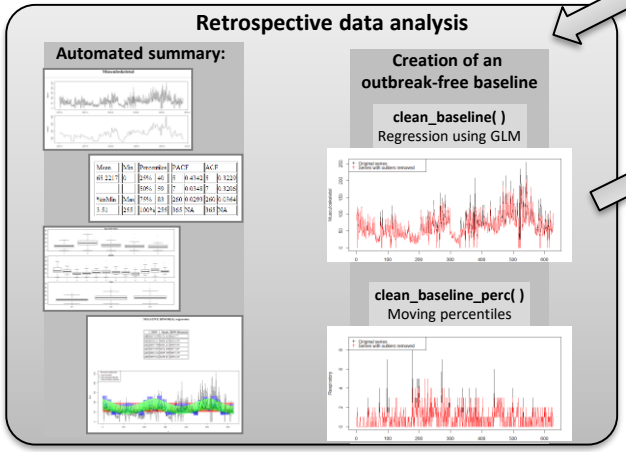
Discussion

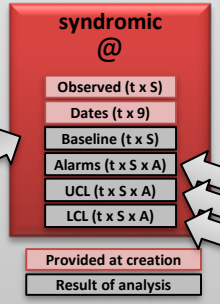
R packages are available which support specific data analysis algorithms, but their application requires experienced users capable of incorporating these algorithms into the larger framework of a syndromic surveillance system. We propose a package in which various statistical analyses already tested for use in animal data streams will be available, coupled with data management capabilities and output visualization.

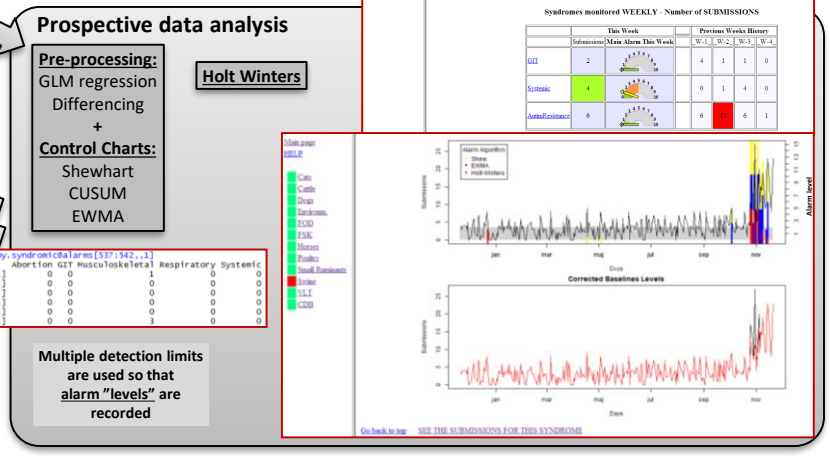
The main innovation offered by the package will be the ability to manage data streams, analyses, alarms and the user interface in a continuous flow; all contained in a single, open source package. This will facilitate the implementation of syndromic surveillance systems by veterinary epidemiologists.











{vetsyn}

```
> data(lab.daily)
```

	SubmissionID	Syndrome	AnimalID	HerdID	MaterialID	DateofSubmission
124	110103-0051	GIT	135856	06115	000009	03/01/2011 00:00:00 AM CET
125	110103-0051	GIT	135856	06115	000009	03/01/2011 00:00:00 AM CET
126	110103-0051	GIT	135856	06115	000009	03/01/2011 00:00:00 AM CET
1744	110104-0305	Musculoskeletal	124347	00001	000424	04/01/2011 00:00:00 AM CET
1745	110104-0307	Musculoskeletal	165122	00001	000426	04/01/2011 00:00:00 AM CET
1809	110104-0328	GIT	111756	00001	000214	04/01/2011 00:00:00 AM CET
1811	110104-0328	GIT	111756	00001	000214	04/01/2011 00:00:00 AM CET
1814	110104-0328	GIT	111756	00001	000214	04/01/2011 00:00:00 AM CET
1923	110104-0388	Respiratory	190339	00163	001202	04/01/2011 00:00:00 AM CET
1924	110104-0388	Respiratory	191057	00163	001206	04/01/2011 00:00:00 AM CET

Classified, "raw" data

raw_to_syndromic()

Syndromic

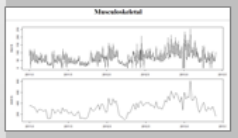
The S4 class center object

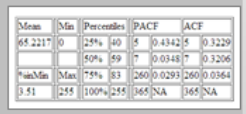
```
> my.syndromic@observed
     Abortion GIT Musculoskeletal Respiratory Systemic
[1,]        0   1              81           0        0
[2,]        0   1             122           2        1
[3,]        2   3              92           1        0
[4,]        0   0               0           0        0
[5,]        1   1              59           0        0
```

```
> my.syndromic@dates
        dates mday month year yday week dow weekday
1  2011-01-03    3     0 2011    2    1   1       1
2  2011-01-04    4     0 2011    3    1   2       1
3  2011-01-05    5     0 2011    4    1   3       1
4  2011-01-06    6     0 2011    5    1   4       1
5  2011-01-07    7     0 2011    6    1   5       1
6  2011-01-08    8     0 2011    7    1   6       0
```

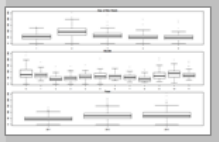
Retrospective data analysis

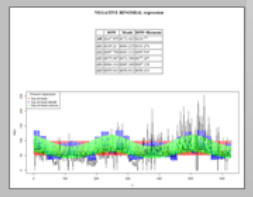
Automated summary:





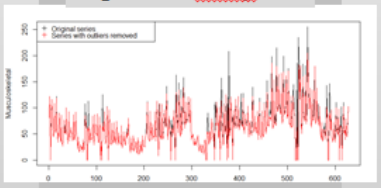
	Mean	Min	Percentiles		PACF		ACF	
65.2217	0	25%	40	5	0.4342	5	0.3229
		50%	59	7	0.0348	7	0.3206
*sinMin	Max	75%	83	260	0.0293	260	0.0364
3.51	255	100%	255	365	NA	365	NA



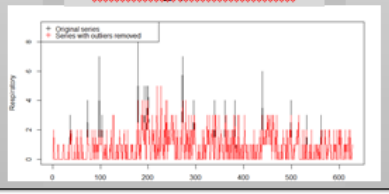


Creation of an outbreak-free baseline

clean_baseline( )
Regression using GLM



clean_baseline_perc( )
Moving percentiles



syndromic @

- Observed (t x S)
- Dates (t x 9)
- Baseline (t x S)
- Alarms (t x S x A)
- UCL (t x S x A)
- LCL (t x S x A)

Provided at creation

Result of analysis

t = # of time points

S = # of syndromic groups

A = # of algorithms used

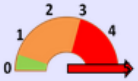
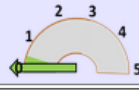
{vetsyn}

Main page
HELP

- Cats
- Cattle
- Dogs
- Environm.
- FOD
- FSK
- Horses
- Poultry
- Small Ruminants
- Swine
- VLT
- CDB

Daily report of syndromes in Swine 2013-11-20

Syndromes monitored DAILY - Number of SUBMISSIONS

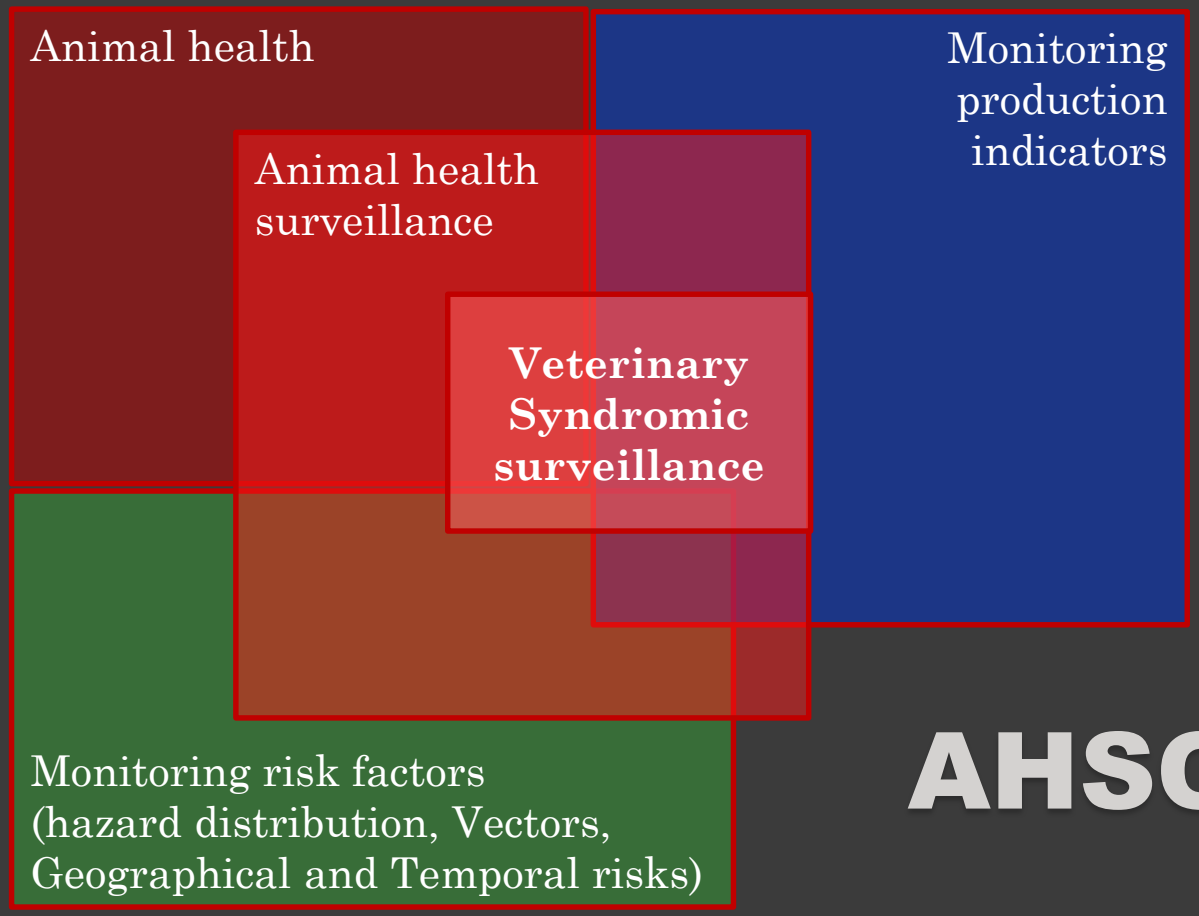
	Today				Previous Days History			
	Submissions	Main Alarm Today	Secondary alarms		_D-1_	_D-2_	_D-3_	_D-4_
Musculoskeletal	132				245	114	95	85
Respiratory	37				2	2	0	4
Fallen Stock	0				1	1	1	0
GIT-Bacterial tests	0				2	0	0	0

Syndromes monitored WEEKLY - Number of SUBMISSIONS

	This Week			Previous Weeks History			
	Submissions	Main Alarm This Week		_W-1_	_W-2_	_W-3_	_W-4_
GIT	2			4	1	1	0
Systemic	4			0	1	4	0
AntimResistance	6			6	18	6	1

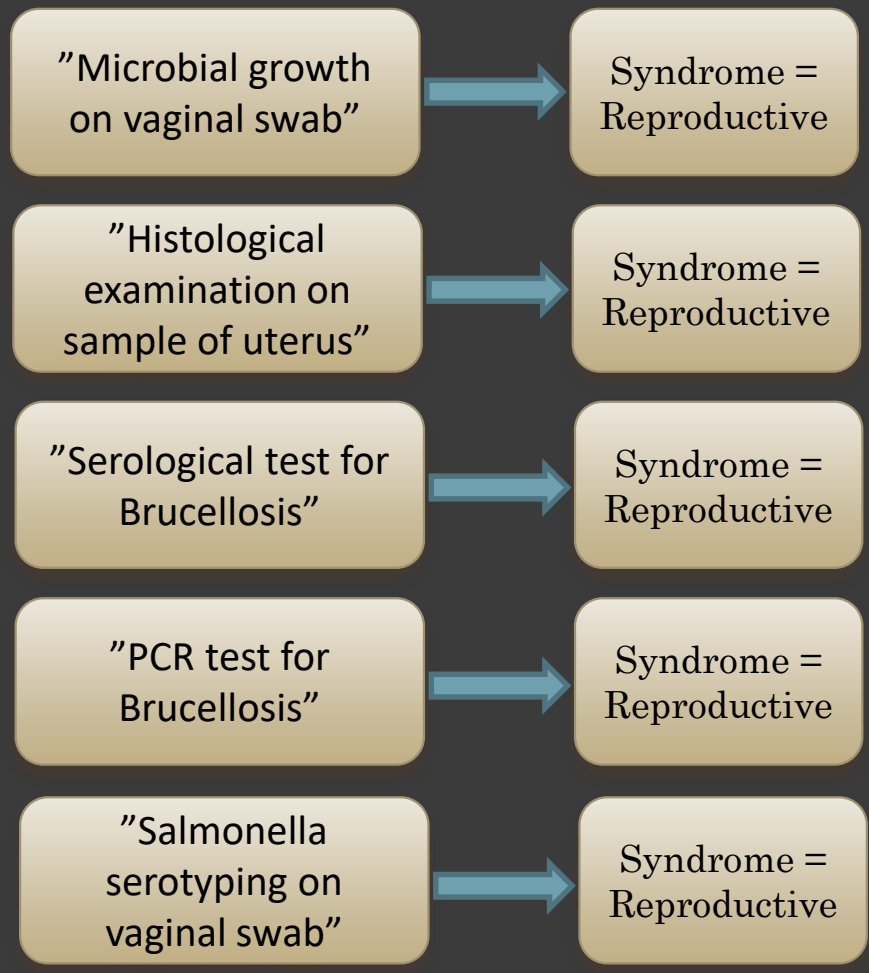
SOFTWARE TOOLS AND RESOURCES

{vetsyn}



AHSO

Animal Health Surveillance Ontology (AHSO)



Consens[illegible] standards for reco[illegible]ata?

We don't need 'smarter' data

But to get the right data to the right place

*So that **smart applications** can do their work*

An **ontology** defines a common vocabulary for those who need to share information in a domain. It includes ***machine-interpretable*** definitions of basic concepts in the domain and relations among them.

- Purpose:
  - to share common understanding of the structure of information among people or software agents
  - to enable reuse of domain knowledge
  - to make domain assumptions explicit
  - to separate domain knowledge from the operational knowledge

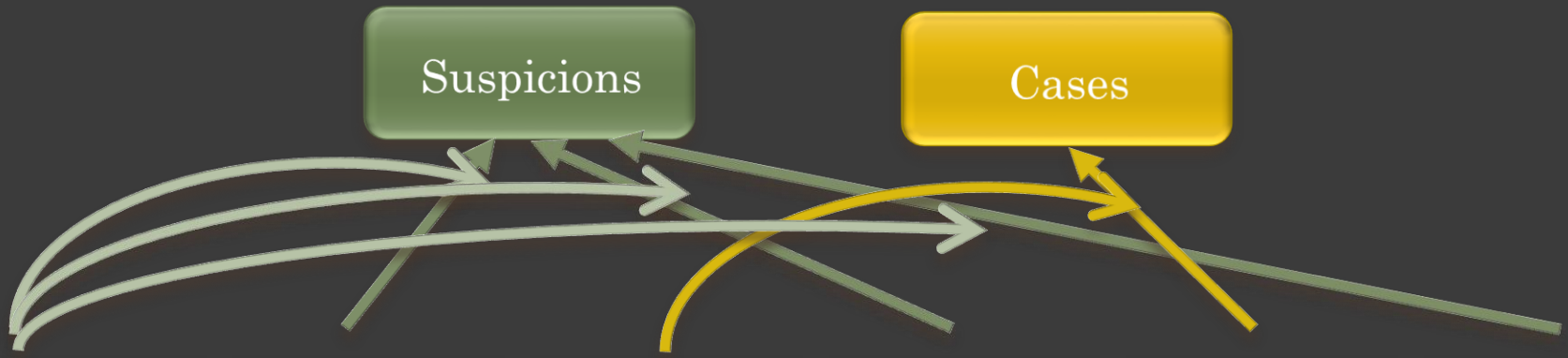
AHSO

Don’t focus on the ‘boxes’…

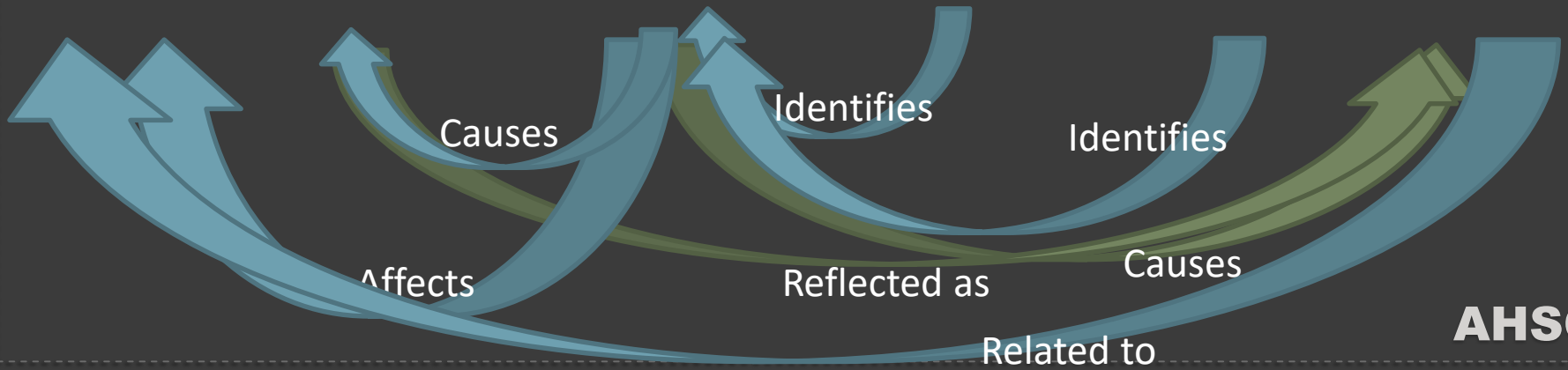
Organ System	Clinical Signs	Threats	Lab Test Ordered	TestResults	Abattoir finding
• Reproductive • Female Genitalia • Vulva • Vagina • Female Internal Reproduct. Organs • Uterus • Ovary • Male External Genitalia • Male Internal Reproductive Organs	• Reproductive • Abortion • Conception failure • OTHER SYNDROMES.. .	• Non-Infectious Threats • Metabolic Disorders • Infectious agents • Brucella sp • Brucella abortus • Brucella suis • Brucella canis • Neospora caninum • BVD • IBR	• UnspecificEtiology • BacterialGrowth • Histology • SerologicalTests • Brucellosis • BVDV • PathogenIdentification • Brucellosis • BVDV • Salmonella	• Serology • Agent Identification • Serological Result	• Ante-Mortem Inspection • Post-Mortem Inspection • Partial Condemnations • Total condemnations

AHSO

But on the rules/relationships…



Organ System	Clinical Signs	Threats	LabTestOrdered	TestResults	Abattoir finding
• Reproductive • Female Genitalia • Vulva • Vagina • Female Internal Reproducitive Organs • Uterus • Ovary • Male External Genitalia • Male Internal Reproductive Organs • OTHER SYSTEMS	• Reproductive • Abortion • Conception failure • OTHER SYNDROMES…	• Non-Infectious Threats • Metabolic Disorders • Infectious agents • Brucella sp • Brucella abortus • Brucella suis • Brucella canis • Neospora caninum • BVD • IBR	• UnspecificEtiology • BacterialGrowth • Histology • SerologicalTests • Brucellosis • BVDV • PathogenIdentification • Brucellosis • BVDV • Salmonella	• Serology • Agent Identification • Serological Result	• Ante-Mortem Inspection • Post-Mortem Inspection • Partial Condemnations • Total condemnations



AHSO

GOOGLE FLU TRENDS

nature International weekly journal of science

Search this journal

ract

er

457, 1012-1014 (19 February 2009) | doi:10.1038/nature07634; Received 14 August
Accepted 13 November 2008; Published online 19 November 2008; Corrected 19
ıry 2009

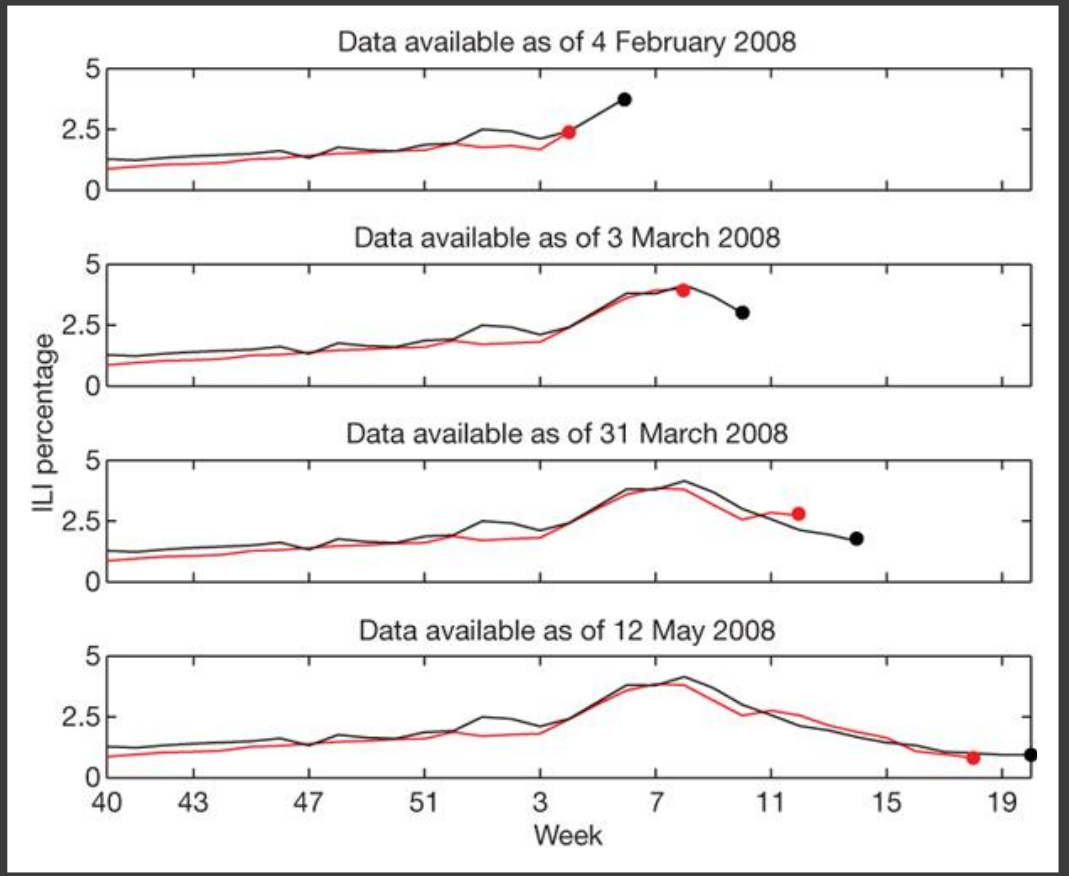
ecting influenza epidemics using search engine
ry data

ıy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2],
S. Smolinski[1] & Larry Brilliant[1]

Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA
Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia
30333, USA

GOOGLE FLU TRENDS

ILI percentages estimated by the model (black) and provided by the CDC (red) showing data available at four points in the 2007-2008 influenza season.



The Economist explains

The backlash against big data

The New York Times

The Opinion Pages | OP-ED CONTRIBUTORS

Eigh

By GARY M

Email

Share

Tweet

Save

More

TRUE STOR
APRIL 17
WATCH TRAILE

alm

FT Magazine

Subscribe for access to FT.c

Home UK World Companies Markets Global Economy Lex Comment Management Person

Arts Magazine Food & Drink House & Home Lunch with the FT Style Books Pursuits Travel Columns How To S

March 28, 2014 11:38 am

Share

Big data: are we making a big mistake?

By Tim Harford

Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media

EDITOR'S CHOICE

SIMON KUPER

A very cosmopolitan spy

GOOGLE FLU TRENDS

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

David Lazer,[1,2]* Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can predict *x* has become commonplace (*5*–*7*) and is often put in sharp contrast

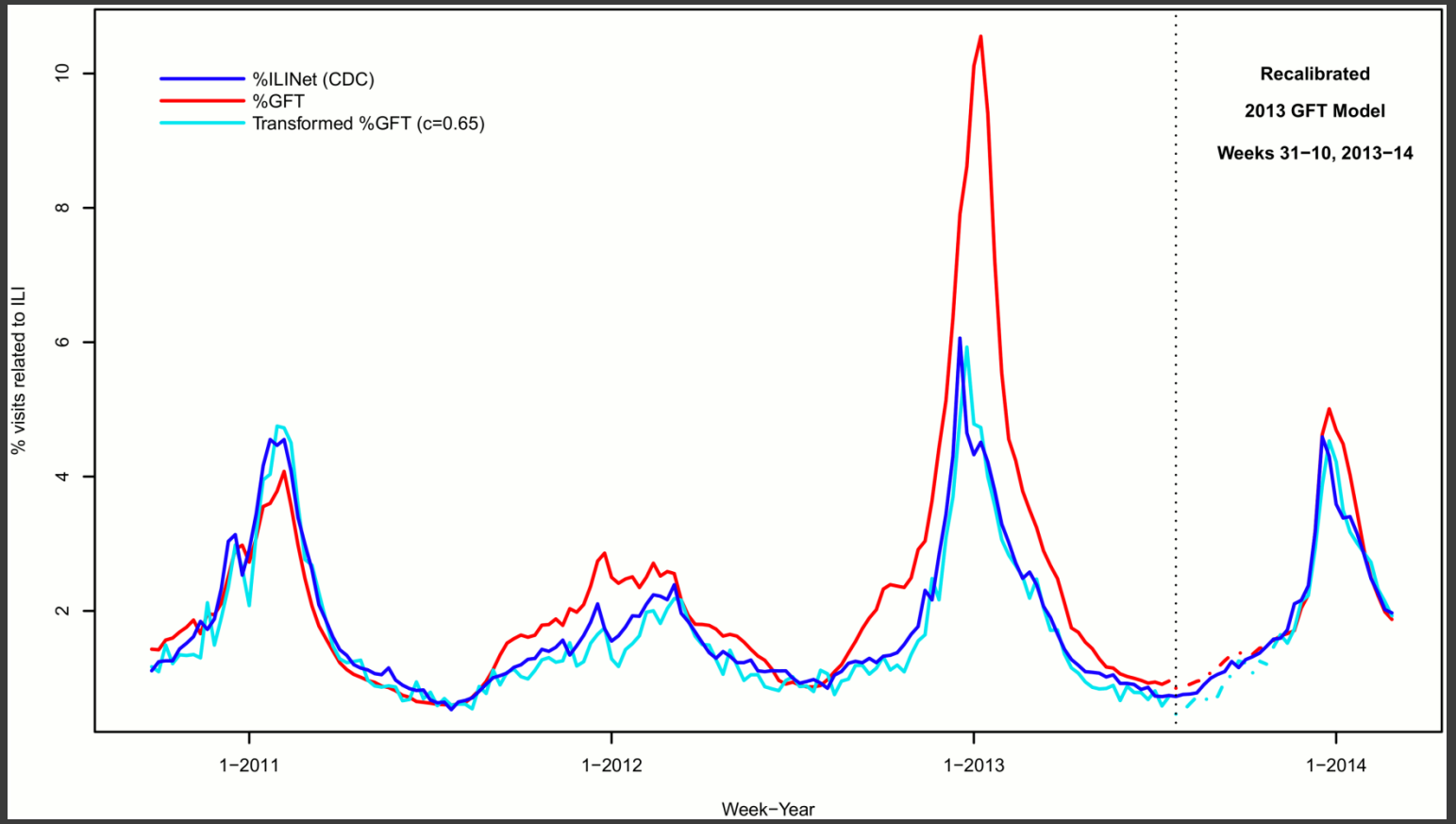
surement and construct validity and reli-

the algorithm in 2009, and this model has run ever since, with a few changes announced in October 2013 (*10*, *15*).

Although not widely reported until 2013, the new GFT has been persistently overestimating flu prevalence for a much longer time. GFT also missed by a very large margin in the 2011–2012 flu season and has missed high for 100 out of 108 weeks starting with August 2011 (see the graph). These errors are not randomly distributed. For example, last week's errors predict this week's errors (temporal autocorrelation), and the direction and magnitude of error varies with the time of year (seasonality). These patterns mean that GFT overlooks considerable information that could be extracted by traditional statistical methods.

Even after GFT was updated in 2009,

GOOGLE FLU TRENDS

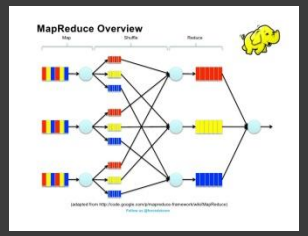


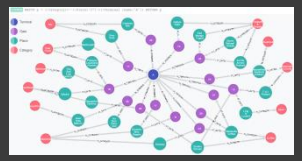
Martin et al (2014) Improving Google Flu Trends Estimates for the United States through Transformation. *PLoS ONE*

Can the use of Lab Data be Effective…? (Even 'unreasonably' so!)

- Large collections of data will continue to become increasingly important
  - Semantic Web / Linked Open Data

- Appropriate methods will be refined:
  - novel Regression and Classification approaches
  - Cluster detection / Visualisation
  - Time series / Aberration detection
  - Structured discovery in Bayesian Networks





- **Fernanda Dórea (SVA)**
- Javier Sanchez (UPEI)
- Ann Lindberg (SVA)
- Beverly McEwen (AHL)
- Bruce McNab (OMAFRA)
- Ann Muckle (UPEI)

ACKNOWLEDGEMENTS